```
OM of: US-08-482-402A-3_COPY_1_848 to: N_Geneseq_1101:*   out_format : pfs

Date: Nov 26, 2001 11:27 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US08482402/runat_26112001_091652_29092/app_query.fasta_1.931
 -DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US08482402_@CGN1_1_272 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-482-402A-3_COPY_1_848
Query length: 848
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 116.270000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation	..
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37493	+	4553.00	5413.30	1.9e-293	3072	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ53061	+	4528.00	5383.56	8.6e-292	3048	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV32403	+	4525.00	5380.62	1.3e-291	2847	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ90304	+	4523.00	5377.59	1.9e-291	3050	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX37301	+	4495.00	5345.90	1.1e-289	2546	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ40728	+	4485.50	5332.88	5.7e-289	3045	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT66437	+	1575.50	1861.91	1.2e-95	3215	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ03118	+	1559.50	1846.15	9.4e-95	2260	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45456	+	1553.50	1838.94	2.4e-94	2272	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF20923	+	1503.00	1777.60	6.2e-91	2558	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA34801	+	1503.00	1777.60	6.2e-91	2558	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21441	+	1503.00	1769.42	1.8e-90	6103	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21436	+	1451.50	1691.47	3.9e-86	35384			
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51671	+	1421.00	1672.59	4.4e-85	5510	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV99922	+	1421.00	1670.54	5.7e-85	6847	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ11842	+	1404.00	1658.99	2.5e-84	2710	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11507	+	1133.50	1342.70	1.0e-66	1386	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD08330	-	875.50	1028.61	3.2e-49	2736	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC93446	+	856.50	1006.30	5.6e-48	2637	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ11843	+	845.50	999.15	1.4e-47	1399	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11509	+	594.50	706.29	2.9e-31	702	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11514	+	561.50	668.18	3.8e-29	615	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11508	+	561.00	666.38	4.8e-29	699	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV72576	-	557.00	669.08	3.4e-29	316	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAD00696	+	425.50	485.39	5.8e-19	5494	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF80223	+	407.50	463.11	1.0e-17	5990	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11515	+	399.50	478.05	1.5e-18	444	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11511	+	361.00	432.72	5.0e-16	417	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11512	+	342.50	411.36	7.8e-15	387	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11516	+	286.50	346.25	3.3e-11	324	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV72575	+	259.00	319.58	1.0e-09	169	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11510	+	255.50	307.46	4.8e-09	393	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI25101	+	224.00	273.89	3.5e-07	257	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI50972	+	224.00	273.89	3.5e-07	257	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11521	+	220.00	269.88	5.9e-07	237	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV23494	+	202.50	208.48	0.0016	17612			
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA13905	+	202.50	208.48	0.0016	17612			
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF30870	+	202.50	208.48	0.0016	17612			
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI15884	+	202.00	241.97	2.1e-05	470	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI37806	+	202.00	241.97	2.1e-05	470	!		
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV11513	+	201.00	246.32	1.2e-05	261			
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51929	+	197.00	216.42	0.0006	3771			
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51928	+	196.00	215.60	0.0006	3624			
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV23481	+	195.00	217.03	0.0005	2742			
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA13898	+	195.00	217.03	0.0005	2742			

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37493

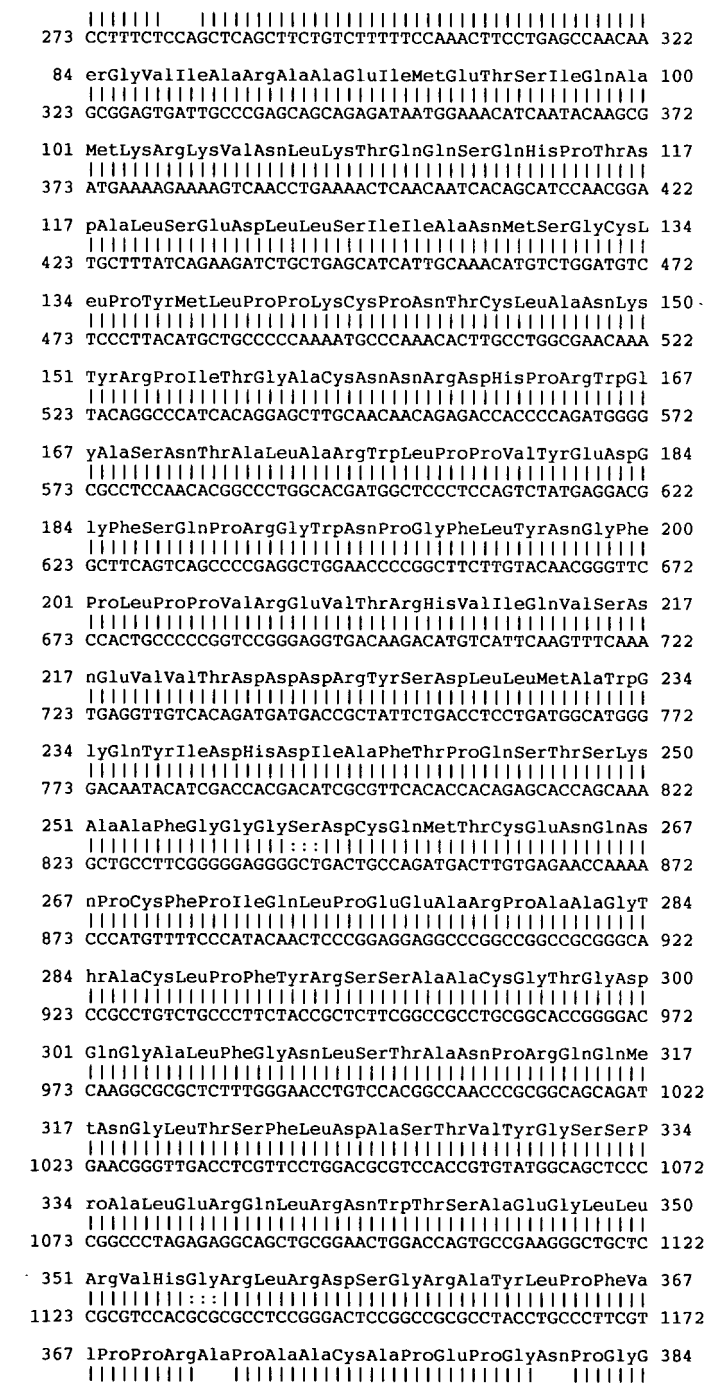
seq_documentation_block:

```
ID   AAQ37493 standard; DNA; 3072 BP.
XX
AC   AAQ37493;
XX
DT   17-JUN-1993  (first entry)
XX
DE   Human TPO gene.
XX
KW   Disease associated B-cell epitope; human thyroid peroxidase;
KW   diagnosis; immune diseases; Hashimoto's thyroiditis;
KW   pHTPO-BS; pHTPO(M1)-BS; site-directed mutagenesis; mutation;
KW   stop codon; EcoRI site; transmembrane region; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             85..2886
FT                   /*tag=  a
FT                   /label= TPO
FT   misc_feature    2631
FT                   /*tag=  b
FT                   /note= "base to be mutated (G -> A) to introduce
FT                        stop codon and EcoRI site"
FT   misc_feature    2634
FT                   /*tag=  c
FT                   /note= "base to be mutated (C -> T) to introduce
FT                        EcoRI site"
FT   misc_feature    2635
FT                   /*tag=  d
FT                   /note= "base to be mutated (T -> C) to introduce
FT                        EcoRI site"
FT   misc_feature    2642
FT                   /*tag=  e
FT                   /note= "base to be mutated (C -> A) to introduce
FT                        stop codon"
XX
PN   WO9303146-A.
XX
PD   18-FEB-1993.
XX
PF   30-JUL-1992;   92WO-US06283.
XX
PR   30-JUL-1991;   91US-0738040.
XX
PA   (RAPO/) RAPOPORT B.
XX
PI   Rapoport B;
XX
DR   WPI; 1993-076503/09.
DR   P-PSDB; AAR32875.
XX
PT   Peptide comprising disease associated B-cell epitope(s) of human
PT   thyroid peroxidase - used for diagnosis and treatment of immune
PT   diseases e.g. Hashimoto's thyroiditis
XX
PS   Disclosure; Fig 7; 131pp; English.
XX
CC   An original full length human TPO cDNA clone in Bluescript (pHTPO-BS)
CC   was submitted to site-directed mutagenesis to produce plasmid pHTPO
CC   (M1)-BS.  The mutations incorporated two stop codons, as well as an
CC   EcoRI site for confirmation, in the region immediately upstream from
CC   the transmembrane region of the human TPO gene.  The entire full
CC   length human TPO gene sequence is given in AAQ37493 for comparison.
CC   As a consequence of the mutation, a "truncated" human TPO protein is
```

```
CC   expressed which is secreted by the host cell rather than bound to is
CC   membrane.
XX
SQ   Sequence 3072 BP; 689 A; 928 C; 890 G; 565 T; 0 other;


alignment_scores:
           Quality: 4553.00              Length:    848
             Ratio:  5.382                Gaps:      0
 Percent Similarity:  99.764   Percent Identity: 99.646

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAQ37493   ..

 Align seg 1/1  to: AAQ37493  from: 1  to: 3072

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 134

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     135 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 184

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 234

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     235 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 284

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     285 CCTTTCTGGAGCTCAGGTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 334

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     335 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 384

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     385 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 434

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     435 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 484

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     485 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 534

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 584

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     585 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 634

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     635 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 684

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     685 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 734

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     735 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 784

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     785 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 834

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     835 GCTGCCTTCGGGGGAGGGTCTGACTGCCAGATGACTTGTGAGAACCAAAA 884

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     885 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 934

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     935 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 984

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     985 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGAGGCAGCAGAT 1034

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1035 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1084

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1085 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1134

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1135 CGCGTCCACGGCCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1184

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1185 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCAACCCCGGAG 1234

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1235 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1284

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1285 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1334

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1335 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1384

     434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1385 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1434

     451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1435 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1484

     467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1485 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1534

     484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1535 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1584

     501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1585 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1634

     517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1635 CTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1684

     534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1685 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1734

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1735 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1784

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1785 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1834

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1835 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1884

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1885 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1934

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1935 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1984

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
      ||||||||||||||||||||||||||  ||||||||||||||||||||||
1985 TCTGGCTGGGAGGCTTAGCTGAAAACTCCCTCCCCAGGGCTCGGACAGGG 2034

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2035 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGACGG 2084

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2085 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2134

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2135 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2184

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2185 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2234

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2235 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2284

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2285 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2334

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2335 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2384

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2385 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2434

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2435 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2484

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
      ||||||||||||||||||||||||||  ||||||||||||||||||||||
2485 GCAGACGGTGCCCACCCCCCCTGCCACGCCTGTGCGAGGTGCAGAAACAC 2534

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
      ||||||||||||||||||||||||||||||||||||||||||||||||||
2535 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2584

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
      |||||||||||||||||||||||||||||||||||||||||||
2585 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGTGACT 2628
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ53061

seq_documentation_block:
ID   AAQ53061 standard; mRNA; 3048 BP.
XX
AC   AAQ53061;
XX
DT   02-JUN-1994  (first entry)
XX
DE   Human thyroid peroxidase.
XX
KW   Peroxidase; plasmid phTPO-2.8; epitope fragment; disease diagnosis;
KW   EC-1.11.1.7; ds.
XX
OS   Homo sapiens.
XX
PN   WO9323073-A.
XX
PD   25-NOV-1993.
XX
PF   22-APR-1993;   93WO-US03837.
XX
PR   19-MAY-1992;   92US-0885656.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Baker JR,  Koenig RJ;
XX
DR   WPI; 1993-386217/48.
DR   P-PSDB; AAR44615.
XX
PT   Isolated specific epitopic regions screening for thyroid
PT   peroxidase auto antibody in sample - by using diagnostic reagent.
PT   for auto immune thyroid disease and for immuno therapy of thyroid
PT   disease and thyroid cancer, for cellular immunity
XX
PS   Disclosure; Page 64-67; 99pp; English.
XX
CC   The epitope peptides encoded by the DNA may be used for
CC   immunotherapy of thyroid diease (e.g. Hashimoto disease) and thyroid
CC   cancer
XX
SQ   Sequence 3048 BP; 671 A; 930 C; 881 G; 566 T; 0 other;


alignment_scores:
           Quality: 4528.00            Length:    848
             Ratio:  5.365              Gaps:      0
 Percent Similarity:  99.528   Percent Identity: 99.175

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAQ53061   ..

 Align seg 1/1  to: AAQ53061  from: 1  to: 3048

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      73 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 122

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     123 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172

      34 roGluGluSerArgValSerServalLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     173 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 222

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
```


273 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 322
84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
323 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372
101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
373 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422
117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472
134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
473 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 522
151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
523 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 572
167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
573 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 622
184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
623 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672
201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722
217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
723 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 772
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822
251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
823 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 872
267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
873 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 922
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
923 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 972
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
973 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1072
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1073 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122
351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1123 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1172
367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1173 GCCGCCACGGCGGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1222
384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1223 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1272
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1273 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1322
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
1323 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1372
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
1373 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1422
451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1423 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1472
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1473 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1522
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
1523 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1572
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1622
517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
1623 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1672
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
1673 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1722
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
1723 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1772
567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
1773 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1822
584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
1823 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1872
601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
1873 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1922
617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1923 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1972
634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
1973 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 2022
651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
2023 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2072
667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
2073 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2122

```
     684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2123 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2172

     701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2173 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2222

     717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
         ||||||||||||||||||||   ||||||||||||||||||||||||||
    2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2272

     734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2273 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2322

     751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2323 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2372

     767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2422

     784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2423 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2472

     801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2473 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2522

     817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2523 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2572

     834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
         |||||||||||||||||||||||||||||||||||:::|||
    2573 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2616
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV32403

seq_documentation_block:
ID   AAV32403 standard; cDNA; 2847 BP.
XX
AC   AAV32403;
XX
DT   25-SEP-1998  (first entry)
XX
DE   Thyroid peroxidase gene.
XX
KW   ds; human; thyroid peroxidase; autoantibody; autoimmune thyroid disease;
KW   Grave's disease; Hashimoto's thyroiditis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2847
FT                   /*tag=  a
FT                   /product= "Thyroid peroxidase"
FT                   /note= "No start codon given"
XX
PN   WO9820354-A2.
XX
PD   14-MAY-1998.
XX
PF   03-NOV-1997;   97WO-GB03014.
XX
PR   01-NOV-1996;   96GB-0022772.
XX
PA   (RSRR-) RSR LTD.
XX
PI   Furmaniak J,  Grennan Jones F,  Rees Smith B;
XX
DR   WPI; 1998-287128/25.
DR   P-PSDB; AAW48781.
XX
PT   Monitoring reactivity of thyroid peroxidase auto-antibodies - is
PT   used to diagnose auto-immune thyroid diseases, or other auto-immune
PT   diseases
XX
PS   Disclosure; Fig 2; 34pp; English.
XX
CC   The thyroid peroxidase (TPO) gene was mutated, to contain truncations or
CC   deletions, to express modified TPO.  The modified TPO was labelled to
CC   monitor the reactivity of TPO autoantibodies (AAb) against TPO AAb in
CC   body fluid from a patient.  The method can be used to diagnose autoimmune
CC   thyroid disease (ATID), especially Grave's disease or Hashimoto's
CC   thyroiditis, or other autoimmune diseases.
XX
SQ   Sequence 2847 BP; 610 A; 885 C; 832 G; 520 T; 0 other;


alignment_scores:
           Quality: 4525.00              Length:    848
             Ratio:  5.355                Gaps:      0
 Percent Similarity:  99.646   Percent Identity: 99.057

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAV32403   ..

 Align seg 1/1  to: AAV32403  from: 1  to: 2847

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      46 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 95

      17 uAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      96 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 145

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     146 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 195

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     196 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 245

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         |||||||   ||||||||||||||||||||||||||||||||||||||||
     246 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 295

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     296 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 345

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     346 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 395

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     396 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 445

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     446 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 495

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     496 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 545

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 546 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 595

 184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 596 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 645

 201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 646 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 695

 217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 696 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 745

 234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 746 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 795

 251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
     ||||||||||||||||||:::||||||||||||||||||||||||||||||
 796 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 845

 267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 846 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 895

 284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 896 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 945

 301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 946 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 995

 317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 996 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1045

 334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1046 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTCCTC 1095

 351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
     |||||||||:::||||||||||||||||||||||||||||||||||||||||
1096 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1145

 367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
     ||||||||||||||||||||||||||||||||||||||   |||||||||
1146 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1195

 384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1196 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1245

 401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1246 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1295

 417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1296 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1345

 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1346 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1395

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1396 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1445

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1446 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1495

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1496 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1545

 501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1546 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1595

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1596 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1645

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1646 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1695

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1696 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1745

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||:::||||||||||||||||||||||||||||||
1746 GTCCAATTCCAGCACCTTGAATCTGGCGTCCATCAACCTGCAGAGGGGCC 1795

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1796 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1845

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1846 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1895

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1896 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1945

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1946 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1995

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1996 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2045

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2046 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2095

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2096 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2145

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2146 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2195

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||   ||||||||||||||||||||||||||||
2196 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2245

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||:::|||||||
2246 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCATGGAGAAT 2295

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2296 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2345

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2346 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2395
```

```
     784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2396 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2445

     801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2446 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2495

     817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2496 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2545

     834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
         ||||||||||||||||||||||||||||||||||:::|||
    2546 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2589

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ90304

seq_documentation_block:
ID   AAQ90304 standard; cDNA to mRNA; 3050 BP.
XX
AC   AAQ90304;
XX
DT   23-OCT-1995  (first entry)
XX
DE   Human thryoid peroxidase gene.
XX
KW   thyroid peroxidase gene; human; recombinant expression; reagent;
KW   antimicrosomal antibody; detection; autoimmune thyroid disease;
KW   myasthenia gravis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..2874
FT                   /*tag=  a
XX
PN   EP655502-A.
XX
PD   31-MAY-1995.
XX
PF   05-SEP-1990;   90EP-0117445.
XX
PR   10-MAY-1990;   90JP-0118770.
PR   05-SEP-1989;   89JP-0228334.
XX
PA   (NIUS ) NIPPON SUISAN KAISHA LTD.
PA   (NISR ) NISSUI PHARM CO LTD.
XX
PI   Hata J,  Kabeno S,  Kato H,  Matsumoto K,  Yagihashi S;
PI   Yamashita S;
XX
DR   WPI; 1995-195590/26.
DR   P-PSDB; AAR75689.
XX
PT   Plasmid for the expression of human thyroid peroxidase - in a form
PT   secreted by Chinese hamster ovary cells.
XX
PS   Disclosure; Page 18-21; 32pp; English.
XX
CC   This nucleotide sequence encodes human thyroid peroxidase (hTPO) and
CC   is used in the formation of plasmid expression vectors suitable for
CC   production of the enzyme in a secretable form from chinese hamster
CC   ovary cells (CHO). The recombinant hTPO produced by the transformed CHO
CC   cells is useful as an assay reagent for the determination of
CC   antimicrosomal antibody levels, esp. for diagnosis of autoimmune
CC   thyroid disease and possibly other diseases, eg. myasthenia gravis,
CC   etc..
XX
SQ   Sequence 3050 BP; 672 A; 931 C; 880 G; 566 T; 1 other;

alignment_scores:
            Quality: 4523.00              Length:    848
              Ratio:  5.359                Gaps:      0
 Percent Similarity:  99.528   Percent Identity: 99.057

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAQ90304   ..

 Align seg 1/1  to: AAQ90304  from: 1  to: 3050

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      73 ATGAGAGCGCTCGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 122

      17 uAlaPheProProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     123 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     173 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 222

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         |||||||   ||||||||||||||||||||||||||||||||||||||||
     273 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 322

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     323 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     373 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 522

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     523 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 572

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 622

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     723 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 772

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||:::||||||||||||||||||||||||||||
     823 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 872
```

```
 267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 873 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 922

 284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 923 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 972

 301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 973 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022

 317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1072

 334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1073 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122

 351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
     |||||||||:::||||||||||||||||||||||||||||||||||||||||
1123 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1172

 367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
     |||||||||   |||||||||||||||||||||||||||   |||||||||
1173 GCCGCCACGGCGGCCTGCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1222

 384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1223 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1272

 401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1273 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1322

 417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1323 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1372

 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1373 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1422

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1423 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1472

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1473 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1522

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1523 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1572

 501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1622

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1623 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1672

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1673 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1722

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1723 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1772

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1773 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1822

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1823 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1872

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1873 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1922

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1923 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1972

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1973 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 2022

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2023 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2072

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2073 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2122

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2123 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2172

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2173 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2222

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||||   ||||||||||||||||||||||||||
2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2272

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2273 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2322

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2323 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2372

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||:::||||||||||||||
2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCACCTCACTTGCACCC 2422

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2423 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2472

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2473 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2522

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2523 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2572

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||||||||||||||||||||||||||||||||||||:::|||
2573 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2616
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX37301

seq_documentation_block:
ID   AAX37301 standard; DNA; 2546 BP.
XX
AC   AAX37301;
```

```
XX
DT   02-JUL-1999  (first entry)
XX
DE   Human thyroid peroxidase hTPO DNA.
XX
KW   Thyroid peroxidase; hTPO; antigen; antibody; anti-human; diagnosis;
KW   thyroid disease; severe myasthenia; lupoid hepatitis;
KW   insulin dependent pediatric diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   JP11094833-A.
XX
PD   09-APR-1999.
XX
PF   19-SEP-1997;   97JP-0273743.
XX
PR   19-SEP-1997;   97JP-0273743.
XX
PA   (SRLS-) SRL KK.
XX
DR   WPI; 1999-291616/25.
DR   P-PSDB; AAY07733.
XX
PT   New antigen for immunological determination of anti-human thyroid
PT   peroxidase antibody - and recombinant human thyroid peroxidase
XX
PS   Claim 7; Page 7-10; 12pp; Japanese.
XX
CC   This invention describes a novel antigen useful for immunological
CC   determination of anti-human thyroid peroxidase (hTPO) antibody with
CC   antigen-antibody reaction. The hTPO antibody is prepared by expression
CC   of nucleic acid(s) prepared by substitution, deletion or addition or
CC   insertion of one or more bases to the gene, particularly with deleted
CC   membrane penetrating area and having molecular wt. of 100 kd, and
CC   expressed in insect cells, optionally having a mannose type sugar chain.
CC   The antigen is useful for the diagnosis of thyroid diseases and severe
CC   myasthenia, lupoid hepatitis, and insulin dependent pediatric diabetes.
XX
SQ   Sequence 2546 BP; 542 A; 804 C; 725 G; 475 T; 0 other;


alignment_scores:
            Quality: 4495.00              Length:    848
              Ratio:  5.338                Gaps:      0
 Percent Similarity:  99.292   Percent Identity: 98.349

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAX37301   ..

 Align seg 1/1  to: AAX37301  from: 1  to: 2546

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 50

      17 uAlaPheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||
      51 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 100

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 150

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 200

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         |||||||   |||||||||||||||||||||||||||||||||||||||||
     201 CCTTTCTCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 250

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 300

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 350

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 400

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 450

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 500

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 550

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 600

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 650

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 700

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 750

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||:::||||||||||||||||||||||||||||
     751 GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 800

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 850

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGTAC 900

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 950

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1000

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1050

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         |||||||||:::||||||||||||||||||||||||||||||||||||||
    1051 CGCGTCCACGCGCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1100

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         ||||||||||||||||:::|||||||||||||||||||   |||||||||
    1101 GCCGCCACGCGCGCCTTCGGCCTGTGCGCCCGAGCCCGGCATCCCCGGAG 1150

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||:::|||||||||
```

```
1151 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCACCGAGGTC 1200

 401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1250

 417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1300

 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
     |||||:::||||||||||||||||||||||||||||||||||||||||||
1301 ACCAAGATGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTT 1350

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
     ||||||||||||||||||||||||   :::|||||||||||||||||||
1351 AGGGATTACATCCCCAGGATCCTGGGATCCCAGGCCTTCCAGCAGTACGT 1400

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1450

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1500

 501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
     |||:::||||||||||||||||||||||||||||||||||||||||||||
1501 GTGAAGAGGCTGGACGCCAGCTTCCAAGAACACCCCGACCTGCCCGGGCT 1550

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1600

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1650

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1700

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1750

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1800

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1850

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATT 1900

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
       ||||||||||||||||||||||||||||||||||||||||||||||||
1901 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1950

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1951 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGACGG 2000

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     |   ||||||||||||||||||||||||||||||||||||||||||||||||
2001 TCACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2050

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2051 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2100

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2101 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2150

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2151 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2200

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2201 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2250

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2251 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2300

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2301 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2350

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2351 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2400

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2401 GCAGACGGTGCCCACCCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC 2450

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2451 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2500

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||||||||||||||||||||||||||||||||||:::|||
2501 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2544
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ40728

seq_documentation_block:

```
ID   AAQ40728 standard; DNA; 3045 BP.
XX
AC   AAQ40728;
XX
DT   17-JUN-1993  (first entry)
XX
DE   Human TPO gene lacking bases 2221-2247.
XX
KW   Disease associated B-cell epitope; human thyroid peroxidase;
KW   diagnosis; immune diseases; Hashimoto's thyroiditis;
KW   mutant; mutation; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             85..2859
FT                   /*tag=  a
FT                   /label= TPO
FT                   /note= "TPO lacking amino acids 713-721"
XX
PN   WO9303146-A.
XX
PD   18-FEB-1993.
XX
PF   30-JUL-1992;   92WO-US06283.
XX
PR   30-JUL-1991;   91US-0738040.
XX
PA   (RAPO/) RAPOPORT B.
XX
PI   Rapoport B;
XX
DR   WPI; 1993-076503/09.
```

```
DR   P-PSDB; AAR35445.
XX
PT   Peptide comprising disease associated B-cell epitope(s) of human
PT   thyroid peroxidase - used for diagnosis and treatment of immune
PT   diseases e.g. Hashimoto's thyroiditis
XX
PS   Disclosure; Page 100 + Fig 7; 131pp; English.
XX
CC   A recombinant TPO wherein amino acids 713-721 have been deleted
CC   or replaced is claimed.  The recombinant TPO may be expressed
CC   in non-thyroidal eukaryotic cells and, like native human TPO,
CC   is enzymatically active, is expressed on the cell surface, and is
CC   not a fusion protein.
XX
SQ   Sequence 3045 BP; 682 A; 922 C; 885 G; 556 T; 0 other;


alignment_scores:
            Quality: 4485.50              Length:    848
              Ratio:  5.359                Gaps:      1
 Percent Similarity:  98.703   Percent Identity: 98.585

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAQ40728   ..

 Align seg 1/1  to: AAQ40728  from: 1  to: 3045

       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 ATGAGAGCGCTGGCTGTGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 134

      17 uAlaPheProPheIleSerArgGlyLysGluLeuLeuTrpGlyLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||
     135 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 184

      34 roGluGluSerArgValSerSerValLeuGluGluSerLysArgLeuVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     185 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG 234

      51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     235 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 284

      67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
         ||||||||||||||::::||||||||||||||||||||||||||||||||
     285 CCTTTCTGGAGCTCAGGTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 334

      84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     335 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 384

     101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     385 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 434

     117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     435 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 484

     134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     485 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 534

     151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 584

     167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     585 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 634

     184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     635 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 684

     201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     685 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 734

     217 nGluValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     735 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 784

     234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     785 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 834

     251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     835 GCTGCCTTCGGGGGAGGGTCTGACTGCCAGATGACTTGTGAGAACCAAAA 884

     267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     885 CCCATGTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGGCCGCGGGCA 934

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     935 CCGCCTGTCTGCCCTTCTACCGCTCTTCGGCCGCCTGCGGCACCGGGGAC 984

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     985 CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGAGGCAGCAGAT 1034

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1035 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1084

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1085 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1134

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1135 CGCGTCCACGGCCGCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1184

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1185 GCCGCCACGCGCGCCTGCGGCCTGTGCGCCCGAGCCCGGCAACCCCGGAG 1234

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1235 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCGCCAGCGAGGTC 1284

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1285 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1334

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1335 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1384

     434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1385 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1434

     451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1435 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1484

     467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1485 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1534

     484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1535 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1584
```

```
 501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1585 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1634

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1635 CTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG 1684

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1685 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1734

 551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1735 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1784

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1785 GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1834

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1835 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1884

 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1885 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAG 1934

 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1935 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1984

 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
     ||||||||||||||||||||||||   ||||||||||||||||||||||
1985 TCTGGCTGGGAGGCTTAGCTGAAAACTCCCTCCCCAGGGCTCGGACAGGG 2034

 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2035 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGACGG 2084

 667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2085 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2134

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2135 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2184

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||
2185 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGC..............  2220

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
                 ||||||||||||||||||||||||||||||||||||
2221 .............GACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2257

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2258 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2307

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2308 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2357

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2358 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2407

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2408 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2457

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||   ||||||||||||||||
2458 GCAGACGGTGCCCACCCCCCCTGCCACGCCTGTGCGAGGTGCAGAAACAC 2507

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2508 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2557

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     |||||||||||||||||||||||||||||||||||||||||||
2558 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGTGACT 2601
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT66437

seq_documentation_block:

```
ID   AAT66437 standard; DNA; 3215 BP.
XX
AC   AAT66437;
XX
DT   27-JUN-1997  (first entry)
XX
DE   Myeloperoxidase coding sequence.
XX
KW   Myeloperoxidase; lectin-binding activity; antibacterial; antiviral;
KW   peroxidase; peripheral blood; bone marrow; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             178..2415
FT                   /*tag= a
FT                   /product= Myeloperoxidase
XX
PN   JP09047286-A.
XX
PD   18-FEB-1997.
XX
PF   08-AUG-1995;   95JP-0222601.
XX
PR   08-AUG-1995;   95JP-0222601.
XX
PA   (SRLS-) SRL KK.
XX
DR   WPI; 1997-186990/17.
DR   P-PSDB; AAW17800.
XX
PT   Myeloperoxidase useful in antibacterial and antiviral agents -
PT   comprises glyco-protein contg. a specified amino acid sequence and
PT   having lectin-combining nature
XX
PS   Disclosure; Page 11-14; 15pp; Japanese.
XX
CC   This sequence encodes myeloperoxidase.  The mature protein portion from
CC   residues 156-745 has a molecular weight measured by SDS-PAGE of 90000
CC   Dalton and an optimum pH of 6.  It has lectin-combining activity and
CC   the ratio of absorbence at 430 nm to that at 280 nm is 0.61.  The
CC   myeloperoxidase has antibacterial activity and antiviral activity.
CC   It also has peroxidase activity.  The new myeloperoxidase has a wide
CC   antibacterial and antiviral spectrum.  The protein was isolated from
CC   the peripheral blood of a bone marrow abnormal formation syndrome
CC   patient.
XX
SQ   Sequence 3215 BP; 652 A; 949 C; 933 G; 681 T; 0 other;
```

```
alignment_scores:
            Quality: 1575.50              Length:    749
              Ratio:  3.018                Gaps:     11
 Percent Similarity:  69.693   Percent Identity: 42.857

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAT66437   ..
```

Align seg 1/1 to: AAT66437 from: 1 to: 3215

```
     587 GlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProArgLe 603
         |||||||||||||||||   ||||||:::|||||||||||||||:::
    1960 GGCCTCCCAGGATACAATGCCTGGAGGCGCTTCTGTGGGCTCCCGCAGCC 2009

     603 uGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSerValAlaA 620
         ||||||   ::::::|||:::|||:::::   :::   :::::::|||
    2010 TGAAACTGTGGGCCAGCTGGGCACGGTGCTGAGGAACCTGAAATTGGCGA 2059

     620 spLysIleLeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeu 636
           |||:::::::::   |||      |||:::||||||||||:::|||:::
    2060 GGAAACTGATGGAGCAGTATGGCACGCCCAACAACATCGACATCTGGATG 2109

     637 GlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProLeuPh 653
         ||||||::::::|||   :::      ::::::|||:::|||||||||::
    2110 GGCGGCGTGTCCGAGCCTCTGAAGCGCAAAGGCCGCGTGGGCCCACTCCT 2159

     653 eAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAspTrpP 670
         :||||||:::||||||   |||::::::   ||||||||||||||   |
    2160 CGCCTGCATCATCGGTACCCAGTTCAGGAAGCTCCGGGATGGTGATCGGT 2209

     670 heTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgArgGluLeu 686
         ||||||||||||||:::   ||||||:::      ||||||:::   |||
    2210 TTTGGTGGGAGAACGAGGGTGTGTTCAGCATGCAGCAGCGACAGGCCCTG 2259

     687 GluLysHisSerLeuSerArgValIleCysAspAsnThrGlyLeuThrAr 703
           :::   ||||||   |||:::||||||||||||||||||:::|||
    2260 GCCCAGATCTCATTGCCCCGGATCATCTGCGACAACACAGGCATCACCAC 2309

     703 gValProMet...AspAlaPheGlnValGlyLysPheProGluAspPheG 719
         |||         :::   |||:::   ::::::::::|||:::||||||
    2310 CGTGTCTAAGAACAACATCTTCATGTCCAACTCATATCCCCGGGACTTTG 2359

     719 luSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgGlu 734
           :::|||:::::::::   ::::::||||||   :::|||||||||
    2360 TCAACTGCAGTACACTTCCTGCATTGAACCTGGCTTCCTGGAGGGAA 2406
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:AAQ03118

seq_documentation_block:

ID AAQ03118 standard; cDNA; 2260 BP.
XX
AC AAQ03118;
XX
DT 22-AUG-1990 (first entry)
XX
DE HindIII/HpaII fragment of pNIV2702 contg. expression cassette for
DE human myeloperoxidase (hMPO).
XX
KW Human myeloperoxidase; enzyme; expression cassette; pNIV2703; pNIV2702;
KW macrophage; cytolytic; bacteriolytic; AIDS therapy; burns treatment.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	9..2245
FT		/*tag= a
FT	misc_feature	1..44
FT		/*tag= b
FT		/note="synthetic oligonucleotide"
FT	misc_feature	2198..2260
FT		/*tag= c
FT		/note="synthetic oligonucleotide"

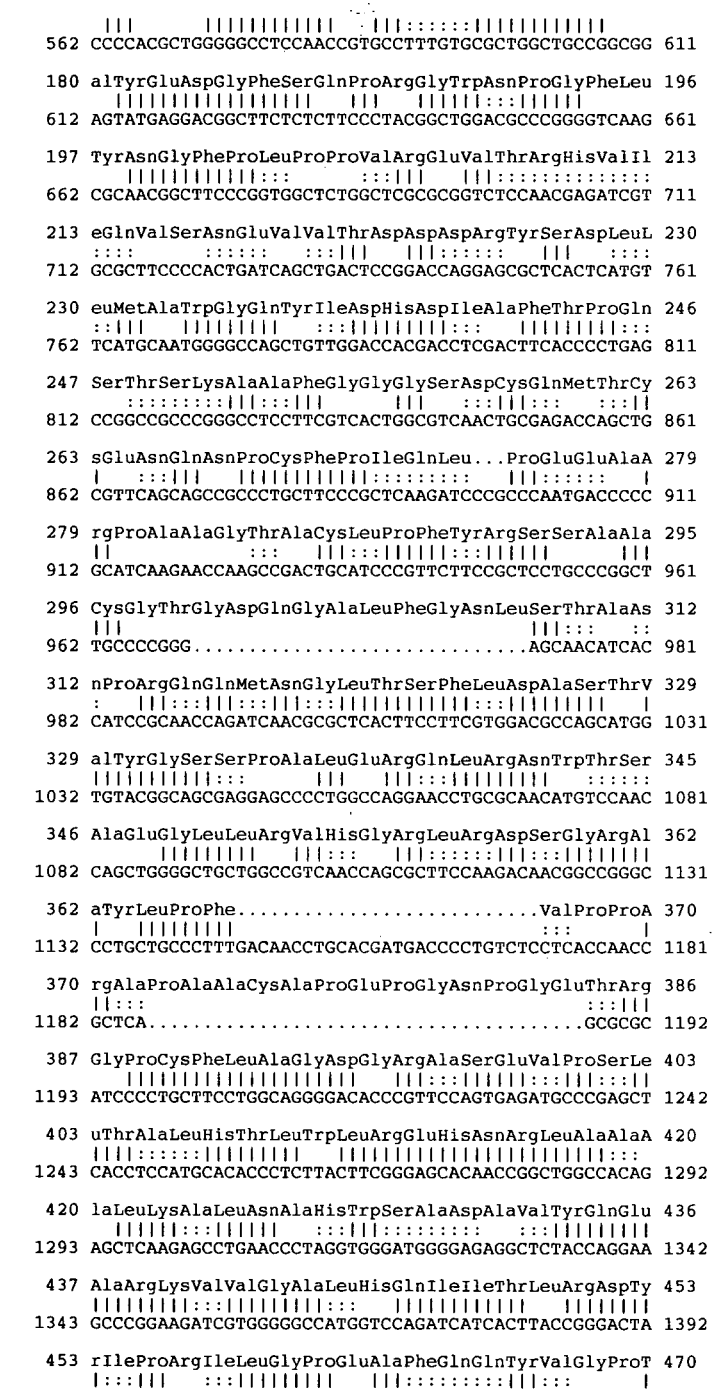
XX
PN WO8912457-A.
XX
PD 28-DEC-1989.
XX
PF 13-JUN-1989; 89WO-EP00668.
XX
PR 14-JUN-1988; 88FR-0007914.
XX
PA (ULBR) UNIV LIBR BRUXELLES.
XX
PI Deby C, Pincemail J, Bollen A;
XX
DR WPI; 1990-022382/03.
XX
PT Human myeloperoxidase prodn. by transformed eukaryotic cells -
PT used to increase cytolytic and bacteriolytic activity in
PT macrophages, eg for treating AIDS or burns
XX
PS Figure 1; ; 45pp; French.
XX
CC It is the complete sequence of hMPO expression cassette. hMPO is produced
CC by culturing prokaryotic or eukaryotic cells transformed with an
CC expression vector for the enzyme. Plasmids pNIV2703 and pNIV2704 are
CC specifically claimed. The former was used for expression in mammalian
CC cells (esp. CHO cells), and the latter was used with wild-type viral DNA
CC to transform cells of Spodoptera frugiperda (esp. Sf9 cells). hMPO can
CC enter macrophages and then increase the cytolytic or bacteriolytic
CC activity of such cells.
XX
SQ Sequence 2260 BP; 440 A; 748 C; 645 G; 427 T; 0 other;

alignment_scores:
Quality: 1559.50 Length: 749
Ratio: 2.982 Gaps: 11
Percent Similarity: 69.826 Percent Identity: 42.857

alignment_block:
US-08-482-402A-3_COPY_1_848 x AAQ03118 ..

Align seg 1/1 to: AAQ03118 from: 1 to: 2260

```
       1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
         |||:::   |||   :::|||:::   :::|||:::::   :::|||
      87 ATGAAGCTGCTTCTGGCCCTAGCAGGCGTCCTGGCCATTCTGGCCACG.. 134

      17 uAlaPhePheProPheIleSerArgGly...LysGluLeuLeuTrpGlyL 33
                 |||      |||:::|||           :::   |||:
     135 .........CCCCAGCCCTCTGAAGGTGCTGCTCCAGCTGTCCTGGGGG 174

      33 ysProGluGluSerArgValSerSerValLeuGluGluSerLysArgLeu 49
         ::   :::   |||   |||   |||   :::||||||:::|||:::|||
     175 AGGTGGACACCTCGTTGGTGCTGAGCTCCATGGAGGAGGCCAAGCAGCTG 224

      50 ValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArg.. 65
         ||||||         |||          :::::::::::|||:::|||
     225 GTGGACAAGGC....TACAAGGAGCGGCGGGAAAGCATCAAGCAGCGGCT 270

      66 .......GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuP 80
               |||      |||      :::|||||||||:::   |||   |
     271 TCGCAGCGGCTCAGCCAGCCCCATGGAACTCCTATCCTACTTCAAGCAGC 320

      80 roGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGluThr 96
         ||      ::::::   :::   :::||||||||||:::   :::::::::
     321 CGGTGGCAGCCACCAGGACGGCGGTGAGGGCCGCTGACTACCTGCACGTG 370

      97 SerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnSerGl 113
         :::::::::   ::::::||||||::::::         ::::::
     371 GCTCTAGACCTGCTGGAGAGGAAGCTGCGGTCCCTGTGGCGAAGGCCATT 420

     113 nHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnM 130
          :::   ||||||:::|||:::         |||:::::::::::::::
     421 CAATGTCACTGATGTGCTGACGCCCGCCCAGCTGAATGTGTTGTCCAAGT 470

     130 etSerGlyCysLeuProTyrMetLeuProProLysCysProAsnThrCys 146
           |||||||||            :::         ||||||:::
     471 CAAGCGGCTGCGCCTACCAGGACGTGGGGGTGACTTGCCCGGAG..... 514

     147 LeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgAspHi 163
               :::|||||||||   |||||||||   ||||||||||||
     515 ...CAGGACAAATACCGCACCATCACCGGGATGTGCAACAACAGACGCAG 561

     163 sProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuProProV 180
```



seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45456

seq_documentation_block:
ID AAZ45456 standard; DNA; 2272 BP.
XX
AC AAZ45456;
XX

```
DT   06-APR-2000  (first entry)
XX
DE   Nucleotide sequence of the human eosinophil peroxidase (EPO).
XX
KW   Human; eosinophil peroxidase; EPO; blood; oxydo-reductase;
KW   respiratory system tissue; ss.
XX
OS   Homo sapiens.
XX
PN   FR2780412-A1.
XX
PD   31-DEC-1999.
XX
PF   30-JUN-1998;   98FR-0008280.
XX
PR   30-JUN-1998;   98FR-0008280.
XX
PA   (HELI-) HELIX BIOTECHNOLOGIES SARL.
XX
PI   Gautier C,  Duport JM;
XX
DR   WPI; 2000-118775/11.
XX
PT   New human recombinant eosinophil peroxidase, useful for replacing human
PT   eosinophil peroxidase -
XX
PS   Claim 3; Fig 1; 29pp; French.
XX
CC   The present sequence represents an expression cassette of a human
CC   eosinophil peroxidase (EPO). The sequence is derived from a
CC   BamHI-HindIII fragment. The specification describes the production of
CC   human EPO fragments by genetic engineering. The sequence was obtained
CC   from total RNA isolated from human blood. The EPO protein is from the
CC   family of oxydo-reductases. The molecular mass of the protein is
CC   about 71 kDa. The human EPO can be used in large scale industrial
CC   applications, especially for the research of new medicaments, e.g. the
CC   toxic effects of natural eosinophil peroxidase on the tissues of the
CC   respiratory system could be diminished or modulated by using a
CC   medicament directly acting on the EPO. Antibodies against the EPO
CC   protein can be used for immunodetection of the protein.
XX
SQ   Sequence 2272 BP; 461 A; 732 C; 627 G; 452 T; 0 other;


alignment_scores:
           Quality: 1553.50              Length:    750
             Ratio:  3.005                Gaps:     11
 Percent Similarity:  68.933   Percent Identity: 43.200

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAZ45456   ..

 Align seg 1/1  to: AAZ45456  from: 1  to: 2272

       3 AlaLeuAlaValLeuSerValThrLeuValMetAla......Cys..... 15
         |||||||||   :::   :::||||||||||:::|||      |||
      40 GCCCTGGCAGGGGTCCTGGCCACACTCGTCCTCGCCCAGCCCTGTGAGGG 89

      16 .ThrGluAlaPhePheProPheIleSerArgGlyLysGluLeuLeuTrpG 32
          |||:::         |||                            |
      90 CACTGACCCAGCCTCCCCT............................G 109

      32 lyLysProGluGluSerArgValSerServalLeuGluGluSerLysArg 48
         ||      |||   |||   :::   :::   :::   |||:::|||
     110 GGGCAGTGGAGACCTCGGTCCTGCGAGACTGCATAGCAGAGGCCAAGTTG 159

      49 LeuValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysAr 65
         |||||||||:::|||   |||         |||:::::::::|||:::||
     160 CTGGTGGATGCTGCC...TACAATTGGACCCAGAAGAGCATCAAGCAGCG 206

      65 g.........GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLysL 79
         |         |||      |||      :::||||||||||:::   |||
     207 GCTTCGCAGCGGTTCAGCCAGCCCCATGGACCTCCTGTCCTACTTCAAAC 256

      79 euProGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGlu 95
           |||      ::::::   :::::::::|||||||||:::   |||:::
     257 AACCGGTAGCAGCCACCAGGACAGTTGTTCGGGCCGCAGATTATATGCAT 306

      96 ThrSerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnSe 112
         :::::::::      :::::::::|||::::::   :::   :::
     307 GTGGCTTTGGGGCTGCTTGAAGAGAAGTTACAACCCCAGCGGTCCGGACC 356

     112 rGlnHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAlaA 129
             :::   ||||||:::|||:::|||      |||   ::::::::::
     357 CTTCAATGTCACTGATGTGCTAACAGAACCACACCTGCGGCTGCTGTCCC 406

     129 snMetSerGlyCysLeuProTyrMetLeuProProLysCysProAsnThr 145
         ::   |||||||||                       :::|||
     407 AGGCCAGTGGCTGTGCTCTCCGGGACCAGGCCGAGCGCTGC..... 447

     146 CysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgAs 162
               ::::::|||||||||   |||||||||   |||||||||:::
     448 ......AGCGACAAGTACCGCACCATCACTGGACGGTGCAACAACAAGAG 491

     162 pHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuProP 179
          :::|||      ||||||||||||   ||||||||||||||||||||
     492 GAGACCCTTGCTAGGGGCCTCCAACCAGGCTCTGGCTCGCTGGCTGCCCG 541

     179 roValTyrGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPhe 195
               ||||||||||||:::|||   |||   ||||||:::|||:::
     542 CCGAGTATGAGGATGGGCTGTCGCTCCCCTTCGGCTGGACCCCCAGCAGG 591

     196 LeuTyrAsnGlyPheProLeuProProValArgGluValThrArgHisVa 212
               |||||||||   ||||||   ||||||   |||:::::::::::
     592 AGGCGCAATGGCTTCCTTCTCCCTCTTGTCCGGGCTGTCTCCAACCAGAT 641

     212 lIleGlnValSerAsnGluValValThrAspAspAspArgTyrSerAspL 229
         :::::::      ||||||   :::|||:::|||          :::   :
     642 TGTGCGCTTCCCCAATGAGAGACTGACCTCCGACCGTGGCCGAGCCCTCA 691

     229 euLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrPro 245
         :::::|||   |||||||||:::||||||||||||:::   |||:::|||
     692 TGTTCATGCAGTGGGGCCAGTTCATTGACCATGACCTGGACTTCTCCCCG 741

     246 GlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetTh 262
         :::|||   :::::::::||||||   :::|||   ||||||:::   ||
     742 GAGTCCCCGGCCAGAGTGGCCTTCACTGCAGGCGTTGACTGTGAGAGGAC 791

     262 rCysGluAsnGlnAsnProCysPheProIleGlnLeu...ProGluGluA 278
         ||||   :::      ||||||||||||||:::::   |||::::::
     792 CTGCGCCCAGCTGCCCCCCTGCTTTCCCATCAAGATCCCACCCAATGACC 841

     278 laArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSerSerAla 294
           |||                  |||:::||||||:::||||||:::
     842 CCCGCATCAAGAACCAGCGTGACTGCATCCCTTTCTTCCGCTCGGCACCC 891

     295 AlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrAl 311
         :::|||      :::   :::
     892 TCATGCCCCCAAAACAAGAAC........................... 912

     311 aAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSerT 328
          :::   |||:::|||:::|||:::||||||||||||:::|||||||||
     913 .AGAGTCCGCAACCAGATCAACGCGCTCACCTCCTTTGTGGACGCCAGCA 961

     328 hrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAsnTrpThr 344
           ||||||||||||:::   :::|||:::   :::|||||||||   |||
     962 TGGTGTATGGCAGTGAGGTCTCCCTCTCGCTGCGGCTCCGCAACCGGACC 1011

     345 SerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspSerGlyAr 361
         :::      |||||||||   ::::::   |||:::::|||:::|||||
    1012 AACTACCTGGGGCTGCTGGCCATCAACCAGCGCTTTCAAGACAACGGCCG 1061

     361 gAlaTyrLeuProPhe.........................ValProP 369
         ||||   |||||||||                          :::
    1062 GGCCCTGCTGCCCTTCGACAACCTGCACGATGACCCCTGTCTCCTCACCA 1111
```

```
 369 roArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyGluThr 385
1112 ACCGCTCG.....................................GCG 1122

 386 ArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluValProSe 402
1123 CGCATCCCCTGCTTCCTGGCAGGTGACACCCGATCAACGGAAACCCCCAA 1172

 402 rLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaA 419
1173 ACTGGCAGCCATGCACACCCTCTTTATGCGAGAGCACAACCGGCTGGCCA 1222

 419 laAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGln 435
1223 CCGAGCTGAGACGCCTGAATCCCCGGTGGAATGGAGACAAACTGTACAAT 1272

 436 GluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeuArgAs 452
1273 GAGGCTCGGAAGATCATGGGGGCCATGGTCCAGATCATCACCTACCGAGA 1322

 452 pTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrValGlyP 469
1323 CTTTCTGCCCCTGGTTCTGGGCAAGGCCCGGGCCAGGAGAGCCCTGGGGC 1372

 469 roTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnValPhe 485
1373 ACTACAGGGGGTACTGCTCCAATGTGGACCCACGGGTGGCCAATGTCTTC 1422

 486 SerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeuValAr 502
1423 ...ACCCTGGCCTTCCGCTTTGGCCACACAATGCTCCAGCCCTTCATGTT 1469

 502 gArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLeuTrpL 519
1470 CCGCTTGGACAGTCAGTACCGGGCCTCCGCACCCAACTCGCATGTCCCAC 1519

 519 euHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLeu 535
1520 TTAGCTCTGCCTTCTTTGCCAGCTGGCGGATCGTGTATGAAGGGGGCATC 1569

 536 AspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValGl 552
1570 GACCCCATCCTCCGGGGCCTCATGGCCACCCCTGCCAAGCTGAACCGTCA 1619

 552 nAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLeuSerA 569
1620 GGATGCCATGTTAGTGGATGAGCTCCGGGACCGGCTGTTTCGGCAAGTGA 1669

 569 snSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAsp 585
1670 GGAGGATTGGGCTGGACCTGGCAGCTCTCAACATGCAACGAAGCCGGGAC 1719

 586 HisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProAr 602
1720 CACGGCCTTCCAGGGTACAATGCTTGGAGGCGCTTCTGTGGGCTCTCCCA 1769

 602 gLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSerValA 619
1770 GCCCCGGAATTTGGCACAGCTTAGCCGGGTGCTGAAAAACCAGGACTTGG 1819

 619 laAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspValTrp 635
1820 CAAGGAAGTTCCTGAATTTGTATGGAACACCTGACAACATTGACATCTGG 1869

 636 LeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProLe 652
1870 ATTGGGGCCATCGCTGAGCCTCTTTTGCCGGGGGCTCGAGTGGGGCCTCT 1919

 652 uPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAspT 669
1920 TCTGGCTTGTCTGTTCGAGAACCAGTTCAGAAGAGCCCGAGACGGAGACA 1969

 669 rpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgArgGlu 685

1970 GGTTCTGGTGGCAGAAACGAGGTGTTTTCACCAAAAGACAGCGCAAGGCC 2019

 686 LeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGlyLeuTh 702
2020 CTGAGCAGAATTTCCTTGTCTCGAATTATATGTGACAATACCGGTATCAC 2069

 702 rArgValProMetAspAlaPheGlnValGlyLysPheProGluAspPheG 719
2070 CACGGTTTCAAGGGACATCTTCAGAGCCAACATCTACCCTCGGGGCTTTG 2119

 719 luSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluThr 735
2120 TGAACTGCAGCCGTATCCCCAGGTTGAACCTATCAGCCTGGCGAGGGACA 2169
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF20923

seq_documentation_block:
ID   AAF20923 standard; DNA; 2558 BP.
XX
AC   AAF20923;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human eosinophil peroxidase polynucleotide fragment #2490.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 144-145; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
```

```
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 2558 BP; 527 A; 820 C; 698 G; 513 T; 0 other;


alignment_scores:
           Quality: 1503.00              Length:    718
             Ratio:  2.982                Gaps:     10
 Percent Similarity:  70.195   Percent Identity: 43.733

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAF20923   ..

 Align seg 1/1  to: AAF20923  from: 1  to: 2558

      32 GlyLysProGluGluSerArgValSerSerValLeuGluGluSerLysAr 48
         |||      |||   |||   :::   :::   :::   |||:::|||
      49 GGGGCAGTGGAGACCTCGGTCCTGCGAGACTGCATAGCAGAGGCCAAGTT 98

      48 gLeuValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysA 65
          |||||||||:::|||   |||         |||:::::::::|||:::|
      99 GCTGGTGGATGCTGCC...TACAATTGGACCCAGAAGAGCATCAAGCAGC 145

      65 rg.........GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLys 78
         ||         |||      |||      :::||||||||||:::   |||
     146 GGCTTCGCAGCGGTTCAGCCAGCCCCATGGACCTCCTGTCCTACTTCAAA 195

      79 LeuProGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGl 95
            |||      ::::::   :::::::::||||||||:::   |||::
     196 CAACCGGTAGCAGCCACCAGGACAGTTGTTCGGGCCGCAGATTATATGCA 245

      95 uThrSerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnS 112
         ::::::::::      :::::::::|||::::::   :::   :::
     246 TGTGGCTTTGGGGCTGCTTGAAGAGAAGTTACAACCCCAGCGGTCCGGAC 295

     112 erGlnHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAla 128
                 ||||||:::|||:::|||      |||   :::::::::
     296 CCTTCATTGTCACTGATGTGCTAACAGAACCACAGCTGCGGCTGCTGTCC 345

     129 AsnMetSerGlyCysLeuProTyrMetLeuProProLysCysProAsnTh 145
         :::   |||||||||                  :::|||
     346 CAGGCCAGTGGCTGTGCTCTCCGGGACCAGGCCGAGCGCTGC........ 387

     145 rCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgA 162
               ::::::|||||||||   |||||||||   |||||||||:::
     388 .......AGCGACAAGTACCGCACCATCACTGGACGGTGCAACAACAAGA 430

     162 spHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuPro 178
           :::|||      ||||||||||||   ||||||||||||||||||||
     431 GGAGACCCTTGCTAGGGGCCTCCAACCAGGCTCTGGCTCGCTGGCTGCCC 480

     179 ProValTyrGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPh 195
               ||||||||||||:::|||   |||   ||||||:::|||:::
     481 GCCGAGTATGAGGATGGGCTGTCGCTCCCCTTCGGCTGGACCCCCAGCAG 530

     195 eLeuTyrAsnGlyPheProLeuProProValArgGluValThrArgHisV 212
               |||||||||   ||||||   ||||||   |||::::::::::
     531 GAGGCGCAATGGCTTCCTTCTCCCTCTTGTCCGGGCTGTCTCCAACCAGA 580

     212 alIleGlnValSerAsnGluValValThrAspAspAspArgTyrSerAsp 228
         ::::::::      ||||||   :::|||:::|||         :::
     581 TTGTGCGCTTCCCCAATGAGAGACTGACCTCCGACCGTGGCCGAGCCCTC 630

     229 LeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrPr 245
         ::::::|||   |||||||||:::||||||||||||:::   |||:::||
     631 ATGTTCATGCAGTGGGGCCAGTTCATTGACCATGACCTGGACTTCTCCCC 680

     245 oGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetT 262
         |:::|||   :::::::::||||||   :::|||   ||||||:::   |
     681 GGAGTCCCCGGCCAGAGTGGCCTTCACTGCAGGCGTTGACTGTGAGAGGA 730

     262 hrCysGluAsnGlnAsnProCysPheProIleGlnLeu...ProGluGlu 277
         |||||   :::      ||||||||||||||||:::::   |||::::::
     731 CCTGCGCCCAGCTGCCCCCCTGCTTTCCCATCAAGATCCCACCCAATGAC 780

     278 AlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSerSerAl 294
            |||                 |||:::||||||:::||||||:::
     781 CCCCGCATCAAGAACCAGCGTGACTGCATCCCTTTCTTCCGCTCGGCACC 830

     294 aAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrA 311
         :::|||      :::   :::
     831 CTCATGCCCCCAAAACAAGAAC..................... 852

     311 laAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSer 327
           ::: ' |||:::|||:::|||:::|||||||||||:::|||||||||
     853 ..AGAGTCCGCAACCAGATCAACGCGCTCACCTCCTTTGTGGACGCCAGC 900

     328 ThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAsnTrpTh 344
            ||||||||||||:::   :::|||:::   :::|||||||||   ||
     901 ATGGTGTATGGCAGTGAGGTCTCCCTCTCGCTGCGGCTCCGCAACCGGAC 950

     344 rSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspSerGlyA 361
         |:::      |||||||||   ::::::   |||:::::::|||:::||||
     951 CAACTACCTGGGGCTGCTGGCCATCAACCAGCGCTTTCAAGACAACGGCC 1000

     361 rgAlaTyrLeuProPhe.........................ValPro 368
         |||||   |||||||||                           :::
    1001 GGGCCCTGCTGCCCTTCGACAACCTGCACGATGACCCCTGTCTCCTCACC 1050

     369 ProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyGluTh 385
            |||:::                                     ::
    1051 AACCGCTCG.....................................GC 1061

     385 rArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluValProS 402
         :|||   ||||||||||||||||||||   |||:::::|||:::|||:
    1062 GCGCATCCCCTGCTTCCTGGCAGGTGACACCCGATCAACGGAAACCCCCA 1111

     402 erLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAla 418
         ::|||:::|||:::|||||||||:::::|||||||||||||||||||||
    1112 AACTGGCAGCCATGCACACCCTCTTTATGCGAGAGCACAACCGGCTGGCC 1161

     419 AlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGl 435
         :::   |||:::   ||||||   :::|||:::::::|||   :::|||::
    1162 ACCGAGCTGAGACGCCTGAATCCCCGGTGGAATGGAGACAAACTGTACAA 1211

     435 nGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeuArgA 452
         :|||||||||||:::::|||||:::   ||||||||||   ||||
    1212 TGAGGCTCGGAAGATCATGGGGGCCATGGTCCAGATCATCACCTACCGAG 1261

     452 spTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrValGly 468
         ||:::::|||   :::||||||            ::::::   :::|||
    1262 ACTTTCTGCCCCTGGTTCTGGGCAAGGCCCGGGCCAGGAGAACCCTGGGG 1311

     469 ProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnValPh 485
            |||:::||||||   |||:::::::::|||   |||:::||||||||
    1312 CACTACAGGGGGTACTGCTCCAATGTGGACCCACGGGTGGCCAATGTCTT 1361

     485 eSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeuValA 502
         |   |||   |||||||||||||||||:::   :::::::|||::::::
    1362 C...ACCCTGGCCTTCCGCTTTGGCCACACAATGCTCCAGCCCTTCATGT 1408
```

```
    502 rgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLeuTrp 518
        |||||||||:::::::::::::                         :::
   1409 TCCGCTTGGACAGTCAGTACCGGGCCTCCGCACCCAACTCGCATGTCCCA 1458

    519 LeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLe 535
        |||   :::||||||||||:::   |||   ::::::      ||||||::
   1459 CTTAGCTCTGCCTTCTTTGCCAGCTGGCGGATCGTGTATGAAGGGGGCAT 1508

    535 uAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValG 552
        :||||||::::::||||||||||:::|||   ||||||||||||:::   |
   1509 CGACCCCATCCTCCGGGGCCTCATGGCCACCCCTGCCAAGCTGAACCGTC 1558

    552 lnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLeuSer 568
        |||||   ::::::   :::|||||||   :::|||||||||
   1559 AGGATGCCATGTTAGTGGATGAGCTCCGGGACCGGCTGTTTCGGCAAGTG 1608

    569 AsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAs 585
        :::         ||||||||||||:::::|||:::||||||:::||||||
   1609 AGGAGGATTGGGCTGGACCTGGCAGCTCTCAACATGCAACGAAGCCGGGA 1658

    585 pHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProA 602
        ||||||||||||||||||||||   ||||||:::||||||||||||   :
   1659 CCACGGCCTTCCAGGGTACAATGCTTGGAGGCGCTTCTGTGGGCTCTCCC 1708

    602 rgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSerVal 618
        ::   ::::::   |||:::|||||   ::::::   ::::::::::::
   1709 AGCCCCGGAATTTGGCACAGCTTAGCCGGGTGCTGAAAAACCAGGACTTG 1758

    619 AlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspValTr 635
        |||   |||:::|||:::||||||      ||||||||||||||:::||
   1759 GCAAGGAAGTTCCTGAATTTGTATGGAACACCTGACAACATTGACATCTG 1808

    635 pLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProL 652
        |:::|||::::::||||||   :::|||||||   ||||||:::|||||||
   1809 GATTGGGGCCATCGCTGAGCCTCTTTTGCCGGGGGCTCGAGTGGGGCCTC 1858

    652 euPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAsp 668
        ||:::||||||||||:::   :::|||::::::      ||||||||||||
   1859 TTCTGGCTTGTCTGTTCGAGAACCAGTTCAGAAGAGC.CGAGACGGAGAC 1907

    669 TrpPheTrpTrpGluAsnSerHisValPheThr.AspAlaGlnArgArgG 685
           |||||||||:::|||   :::||||||::: :::   ||||||:::
   1908 AGGTTCTGGTGGCAGAAC...GAGGTGTTTTCACCAAAGACAGCGCAAGG 1954

    685 luLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGlyLeu 701
          |||::::::   ||||||||||||:::||||||||||||||||||:::
   1955 CCCTGAGCAGAATTTCCTTGTCTCGAATTATATGTGACAATACCGGTATC 2004

    702 ThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAspPh 718
        |||   |||      |||   |||:::::::::   :::|||:::   ||
   2005 ACCACGGTTTCAAGGGACATCTTCAGAGCCAACATCTACCCTCGGGGCTT 2054

    718 eGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluT 735
        |   :::|||:::   |||      :::||||||:::|||||||||   |
   2055 TGTGAACTGCAGCCGTATCCCCAGGTTGAACCTATCAGCCTGGCGAGGGA 2104

    735 hr 735
        ||
   2105 CA 2106

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA34801

seq_documentation_block:
ID   AAA34801 standard; DNA; 2558 BP.
XX
AC   AAA34801;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2490.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 647-648; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 2558 BP; 527 A; 820 C; 698 G; 513 T; 0 other;


alignment_scores:
            Quality: 1503.00              Length:    718
              Ratio:  2.982                Gaps:     10
 Percent Similarity:  70.195   Percent Identity: 43.733

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAA34801   ..

 Align seg 1/1  to: AAA34801  from: 1  to: 2558

      32 GlyLysProGluGluSerArgValSerSerValLeuGluGluSerLysAr 48
         |||      |||   |||   :::   :::   :::   |||:::|||
      49 GGGGCAGTGGAGACCTCGGTCCTGCGAGACTGCATAGCAGAGGCCAAGTT 98
```


48 gLeuValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysA 65
99 GCTGGTGGATGCTGCC...TACAATTGGACCCAGAAGAGCATCAAGCAGC 145
65 rg.....GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLys 78
146 GGCTTCGCAGCGGTTCAGCCAGCCCCATGGACCTCCTGTCCTACTTCAAA 195
79 LeuProGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGl 95
196 CAACCGGTAGCAGCCACCAGGACAGTTGTTCGGGCCGCAGATTATATGCA 245
95 uThrSerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnS 112
246 TGTGGCTTTGGGGCTGCTTGAAGAGAAGTTACAACCCCAGCGGTCCGGAC 295
112 erGlnHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAla 128
296 CCTTCATTGTCACTGATGTGCTAACAGAACCACAGCTGCGGCTGCTGTCC 345
129 AsnMetSerGlyCysLeuProTyrMetLeuProProLysCysProAsnTh 145
346 CAGGCCAGTGGCTGTGCTCTCCGGGACCAGGCCGAGCGCTGC..... 387
145 rCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgA 162
388 .....AGCGACAAGTACCGCACCATCACTGGACGGTGCAACAACAAGA 430
162 spHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuPro 178
431 GGAGACCCTTGCTAGGGGCCTCCAACCAGGCTCTGGCTCGCTGGCTGCCC 480
179 ProValTyrGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPh 195
481 GCCGAGTATGAGGATGGGCTGTCGCTCCCCTTCGGCTGGACCCCCAGCAG 530
195 eLeuTyrAsnGlyPheProLeuProProValArgGluValThrArgHisV 212
531 GAGGCGCAATGGCTTCCTTCTCCCTCTTGTCCGGGCTGTCTCCAACCAGA 580
212 alIleGlnValSerAsnGluValValThrAspAspAspArgTyrSerAsp 228
581 TTGTGCGCTTCCCCAATGAGAGACTGACCTCCGACCGTGGCCGAGCCCTC 630
229 LeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrPr 245
631 ATGTTCATGCAGTGGGGCCAGTTCATTGACCATGACCTGGACTTCTCCCC 680
245 oGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetT 262
681 GGAGTCCCCGGCCAGAGTGGCCTTCACTGCAGGCGTTGACTGTGAGAGGA 730
262 hrCysGluAsnGlnAsnProCysPheProIleGlnLeu...ProGluGlu 277
731 CCTGCGCCCAGCTGCCCCCCTGCTTTCCCATCAAGATCCCACCCAATGAC 780
278 AlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSerSerAl 294
781 CCCCGCATCAAGAACCAGCGTGACTGCATCCCTTTCTTCCGCTCGGCACC 830
294 aAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrA 311
831 CTCATGCCCCCAAAACAAGAAC..... 852
311 laAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSer 327
853 ..AGAGTCCGCAACCAGATCAACGCGCTCACCTCCTTTGTGGACGCCAGC 900
328 ThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAsnTrpTh 344
901 ATGGTGTATGGCAGTGAGGTCTCCCTCTCGCTGCGGCTCCGCAACCGGAC 950
344 rSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspSerGlyA 361
951 CAACTACCTGGGGCTGCTGGCCATCAACCAGCGCTTTCAAGACAACGGCC 1000
361 rgAlaTyrLeuProPhe.....ValPro 368
1001 GGGCCCTGCTGCCCTTCGACAACCTGCACGATGACCCCTGTCTCCTCACC 1050
369 ProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyGluTh 385
1051 AACCGCTCG.....GC 1061
385 rArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluValProS 402
1062 GCGCATCCCCTGCTTCCTGGCAGGTGACACCCGATCAACGGAAACCCCCA 1111
402 erLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAla 418
1112 AACTGGCAGCCATGCACACCCTCTTTATGCGAGAGCACAACCGGCTGGCC 1161
419 AlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGl 435
1162 ACCGAGCTGAGACGCCTGAATCCCCGGTGGAATGGAGACAAACTGTACAA 1211
435 nGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeuArgA 452
1212 TGAGGCTCGGAAGATCATGGGGGCCATGGTCCAGATCATCACCTACCGAG 1261
452 spTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrValGly 468
1262 ACTTTCTGCCCCTGGTTCTGGGCAAGGCCCGGGCCAGGAGAACCCTGGGG 1311
469 ProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnValPh 485
1312 CACTACAGGGGGTACTGCTCCAATGTGGACCCACGGGTGGCCAATGTCTT 1361
485 eSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeuValA 502
1362 C...ACCCTGGCCTTCCGCTTTGGCCACACAATGCTCCAGCCCTTCATGT 1408
502 rgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLeuTrp 518
1409 TCCGCTTGGACAGTCAGTACCGGGCCTCCGCACCCAACTCGCATGTCCCA 1458
519 LeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLe 535
1459 CTTAGCTCTGCCTTCTTTGCCAGCTGGCGGATCGTGTATGAAGGGGGCAT 1508
535 uAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValG 552
1509 CGACCCCATCCTCCGGGGCCTCATGGCCACCCCTGCCAAGCTGAACCGTC 1558
552 lnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLeuSer 568
1559 AGGATGCCATGTTAGTGGATGAGCTCCGGGACCGGCTGTTTCGGCAAGTG 1608
569 AsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAs 585
1609 AGGAGGATTGGGCTGGACCTGGCAGCTCTCAACATGCAACGAAGCCGGGA 1658
585 pHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProA 602
1659 CCACGGCCTTCCAGGGTACAATGCTTGGAGGCGCTTCTGTGGGCTCTCCC 1708
602 rgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSerVal 618
1709 AGCCCCGGAATTTGGCACAGCTTAGCCGGGTGCTGAAAAACCAGGACTTG 1758
619 AlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspValTr 635
1759 GCAAGGAAGTTCCTGAATTTGTATGGAACACCTGACAACATTGACATCTG 1808
635 pLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProL 652

```
    1809 GATTGGGGCCATCGCTGAGCCTCTTTTGCCGGGGGCTCGAGTGGGGCCTC 1858

     652 euPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAsp 668
         ||:::||||||||||:::   :::|||:::::      ||||||||||
    1859 TTCTGGCTTGTCTGTTCGAGAACCAGTTCAGAAGAGC.CGAGACGGAGAC 1907

     669 TrpPheTrpTrpGluAsnSerHisValPheThr.AspAlaGlnArgArgG 685
           |||||||||:::|||   :::||||||::: :::   ||||||:::
    1908 AGGTTCTGGTGGCAGAAC...GAGGTGTTTTCACCAAAGACAGCGCAAGG 1954

     685 luLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGlyLeu 701
           |||::::::   ||||||||||||:::||||||||||||||||||:::
    1955 CCCTGAGCAGAATTTCCTTGTCTCGAATTATATGTGACAATACCGGTATC 2004

     702 ThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAspPh 718
         |||   |||       |||   |||:::::::::   :::|||:::   ||
    2005 ACCACGGTTTCAAGGGACATCTTCAGAGCCAACATCTACCCTCGGGGCTT 2054

     718 eGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluT 735
         |   :::|||:::   |||      :::||||||:::|||||||||   |
    2055 TGTGAACTGCAGCCGTATCCCCAGGTTGAACCTATCAGCCTGGCGAGGGA 2104

     735 hr 735
         ||
    2105 CA 2106
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21441

seq_documentation_block:

ID AAF21441 standard; DNA; 6103 BP.
XX
AC AAF21441;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human eosinophil peroxidase polynucleotide fragment #3008.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 142-143; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6103 BP; 1218 A; 1863 C; 1727 G; 1287 T; 8 other;

alignment_scores:

Quality: 1503.00 Length: 718
Ratio: 2.982 Gaps: 10
Percent Similarity: 70.195 Percent Identity: 43.733

alignment_block:

US-08-482-402A-3_COPY_1_848 x AAF21441 ..

Align seg 1/1 to: AAF21441 from: 1 to: 6103

```
      32 GlyLysProGluGluSerArgValSerSerValLeuGluGluSerLysAr 48
         |||      |||   |||   :::   :::   :::   |||:::|||
    3594 GGGGCAGTGGAGACCTCGGTCCTGCGAGACTGCATAGCAGAGGCCAAGTT 3643

      48 gLeuValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysA 65
          |||||||||:::|||   |||         |||:::::::::|||:::|
    3644 GCTGGTGGATGCTGCC...TACAATTGGACCCAGAAGAGCATCAAGCAGC 3690

      65 rg.........GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLys 78
         ||         |||      |||      :::|||||||||:::   |||
    3691 GGCTTCGCAGCGGTTCAGCCAGCCCCATGGACCTCCTGTCCTACTTCAAA 3740

      79 LeuProGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGl 95
            |||      ::::::   ::::::::::||||||||:::   |||::
    3741 CAACCGGTAGCAGCCACCAGGACAGTTGTTCGGGCCGCAGATTATATGCA 3790

      95 uThrSerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnS 112
         ::::::::::      ::::::::::|||::::::   :::   :::
    3791 TGTGGCTTTGGGGCTGCTTGAAGAGAAGTTACAACCCCAGCGGTCCGGAC 3840

     112 erGlnHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAla 128
                 ||||||:::|||:::|||      |||   :::::::::
    3841 CCTTCATTGTCACTGATGTGCTAACAGAACCACAGCTGCGGCTGCTGTCC 3890

     129 AsnMetSerGlyCysLeuProTyrMetLeuProProLysCysProAsnTh 145
         :::   |||||||||                      :::|||
    3891 CAGGCCAGTGGCTGTGCTCTCCGGGACCAGGCCGAGCGCTGC........ 3932

     145 rCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgA 162
               ::::::||||||||||   |||||||||   ||||||||||:::
    3933 .......AGCGACAAGTACCGCACCATCACTGGACGGTGCAACAACAAGA 3975
```

```
 162 spHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuPro 178
3976 GGAGACCCTTGCTAGGGGCCTCCAACCAGGCTCTGGCTCGCTGGCTGCCC 4025

 179 ProValTyrGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPh 195
4026 GCCGAGTATGAGGATGGGCTGTCGCTCCCCTTCGGCTGGACCCCCAGCAG 4075

 195 eLeuTyrAsnGlyPheProLeuProProValArgGluValThrArgHisV 212
4076 GAGGCGCAATGGCTTCCTTCTCCCTCTTGTCCGGGCTGTCTCCAACCAGA 4125

 212 alIleGlnValSerAsnGluValValThrAspAspAspArgTyrSerAsp 228
4126 TTGTGCGCTTCCCCAATGAGAGACTGACCTCCGACCGTGGCCGAGCCCTC 4175

 229 LeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrPr 245
4176 ATGTTCATGCAGTGGGGCCAGTTCATTGACCATGACCTGGACTTCTCCCC 4225

 245 oGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetT 262
4226 GGAGTCCCCGGCCAGAGTGGCCTTCACTGCAGGCGTTGACTGTGAGAGGA 4275

 262 hrCysGluAsnGlnAsnProCysPheProIleGlnLeu...ProGluGlu 277
4276 CCTGCGCCCAGCTGCCCCCCTGCTTTCCCATCAAGATCCCACCCAATGAC 4325

 278 AlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSerSerAl 294
4326 CCCCGCATCAAGAACCAGCGTGACTGCATCCCTTTCTTCCGCTCGGCACC 4375

 294 aAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrA 311
4376 CTCATGCCCCCAAAACAAGAAC..................................... 4397

 311 laAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSer 327
4398 ..AGAGTCCGCAACCAGATCAACGCGCTCACCTCCTTTGTGGACGCCAGC 4445

 328 ThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAsnTrpTh 344
4446 ATGGTGTATGGCAGTGAGGTCTCCCTCTCGCTGCGGCTCCGCAACCGGAC 4495

 344 rSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspSerGlyA 361
4496 CAACTACCTGGGGCTGCTGGCCATCAACCAGCGCTTTCAAGACAACGGCC 4545

 361 rgAlaTyrLeuProPhe.....................ValPro 368
4546 GGGCCCTGCTGCCCTTCGACAACCTGCACGATGACCCCTGTCTCCTCACC 4595

 369 ProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyGluTh 385
4596 AACCGCTCG.....................................GC 4606

 385 rArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluValProS 402
4607 GCGCATCCCCTGCTTCCTGGCAGGTGACACCCGATCAACGGAAACCCCCA 4656

 402 erLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAla 418
4657 AACTGGCAGCCATGCACACCCTCTTTATGCGAGAGCACAACCGGCTGGCC 4706

 419 AlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGl 435
4707 ACCGAGCTGAGACGCCTGAATCCCCGGTGGAATGGAGACAAACTGTACAA 4756

 435 nGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeuArgA 452
4757 TGAGGCTCGGAAGATCATGGGGGCCATGGTCCAGATCATCACCTACCGAG 4806

 452 spTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrValGly 468
4807 ACTTTCTGCCCCTGGTTCTGGGCAAGGCCCGGGCCAGGAGAACCCTGGGG 4856

 469 ProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnValPh 485
4857 CACTACAGGGGGTACTGCTCCAATGTGGACCCACGGGTGGCCAATGTCTT 4906

 485 eSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeuValA 502
4907 C...ACCCTGGCCTTCCGCTTTGGCCACACAATGCTCCAGCCCTTCATGT 4953

 502 rgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLeuTrp 518
4954 TCCGCTTGGACAGTCAGTACCGGGCCTCCGCACCCAACTCGCATGTCCCA 5003

 519 LeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLe 535
5004 CTTAGCTCTGCCTTCTTTGCCAGCTGGCGGATCGTGTATGAAGGGGGCAT 5053

 535 uAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValG 552
5054 CGACCCCATCCTCCGGGGCCTCATGGCCACCCCTGCCAAGCTGAACCGTC 5103

 552 lnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLeuSer 568
5104 AGGATGCCATGTTAGTGGATGAGCTCCGGGACCGGCTGTTTCGGCAAGTG 5153

 569 AsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyArgAs 585
5154 AGGAGGATTGGGCTGGACCTGGCAGCTCTCAACATGCAACGAAGCCGGGA 5203

 585 pHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeuProA 602
5204 CCACGGCCTTCCAGGGTACAATGCTTGGAGGCGCTTCTGTGGGCTCTCCC 5253

 602 rgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSerVal 618
5254 AGCCCCGGAATTTGGCACAGCTTAGCCGGGTGCTGAAAAACCAGGACTTG 5303

 619 AlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspValTr 635
5304 GCAAGGAAGTTCCTGAATTTGTATGGAACACCTGACAACATTGACATCTG 5353

 635 pLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProL 652
5354 GATTGGGGCCATCGCTGAGCCTCTTTTGCCGGGGGCTCGAGTGGGGCCTC 5403

 652 euPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGlyAsp 668
5404 TTCTGGCTTGTCTGTTCGAGAACCAGTTCAGAAGAGC.CGAGACGGAGAC 5452

 669 TrpPheTrpTrpGluAsnSerHisValPheThr.AspAlaGlnArgArgG 685
5453 AGGTTCTGGTGGCAGAAC...GAGGTGTTTTCACCAAAGACAGCGCAAGG 5499

 685 luLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGlyLeu 701
5500 CCCTGAGCAGAATTTCCTTGTCTCGAATTATATGTGACAATACCGGTATC 5549

 702 ThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAspPh 718
5550 ACCACGGTTTCAAGGGACATCTTCAGAGCCAACATCTACCCTCGGGGCTT 5599

 718 eGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluT 735
5600 TGTGAACTGCAGCCGTATCCCCAGGTTGAACCTATCAGCCTGGCGAGGGA 5649

 735 hr 735
5650 CA 5651

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF21436
```

```
seq_documentation_block:
ID   AAF21436 standard; DNA; 35384 BP.
XX
AC   AAF21436;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human enzyme-related antisense polynucleotide #3003.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 47-55; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX

SQ   Sequence 35384 BP; 7013 A; 10128 C; 10025 G; 7883 T; 335 other;


alignment_scores:
           Quality: 1451.50              Length:    709
             Ratio:  2.950                Gaps:      9
 Percent Similarity:  69.394   Percent Identity: 43.018

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAF21436   ..

 Align seg 1/1  to: AAF21436  from: 1  to: 35384

     32 GlyLysProGluGluSerArgValSerSerValLeuGluGluSerLysAr 48
        |||      |||   |||   :::   :::   :::   |||:::|||
  22141 GGGGCAGTGGAGACCTCGGTCCTGCGAGACTGCATAGCAGAGGCCAAGTT 22190

     48 gLeuValAspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysA 65
         |||||||||:::|||   |||         |||:::::::::|||:::|
  22191 GCTGGTGGATGCTGCC...TACAATTGGACCCAGAAGAGCATCAAGCAGC 22237

     65 rg.........GlyIleLeuSerGlyAlaGlnLeuLeuSerPheSerLys 78
        ||         |||      |||      :::||||||||:::   |||
  22238 GGCTTCGCAGCGGTTCAGCCAGCCCCATGGACCTCCTGTCCTACTTCAAA 22287

     79 LeuProGluProThrSerGlyValIleAlaArgAlaAlaGluIleMetGl 95
           |||      ::::::   :::::::::||||||||:::   |||::
  22288 CAACCGGTAGCAGCCACCAGGACAGTTGTTCGGGCCGCAGATTATATGCA 22337

     95 uThrSerIleGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnS 112
        ::::::::::      ::::::::::|||::::::   :::   :::
  22338 TGTGGCTTTGGGGCTGCTTGAAGAGAAGTTACAACCCCAGCGGTCCGGAC 22387

    112 erGlnHisProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAla 128
                 ||||||:::|||:::|||      |||   :::::::::
  22388 CCTTCATTGTCACTGATGTGCTAACAGAACCACAGCTGCGGCTGCTGTCC 22437

    129 AsnMetSerGlyCysLeuProTyrMetLeuProProLysCysProAsnTh 145
        :::   |||||||||                     :::|||
  22438 CAGGCCAGTGGCTGTGCTCTCCGGGACCAGGCCGAGCGCTGC........ 22479

    145 rCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgA 162
               ::::::|||||||||   |||||||||   |||||||||:::
  22480 .......AGCGACAAGTACCGCACCATCACTGGACGGTGCAACAACAAGA 22522

    162 spHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuPro 178
          :::|||      ||||||||||||   ||||||||||||||||||||
  22523 GGAGACCCTTGCTAGGGGCCTCCAACCAGGCTCTGGCTCGCTGGCTGCCC 22572

    179 ProValTyrGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPh 195
              ||||||||||||:::|||   |||   ||||||:::|||:::
  22573 GCCGAGTATGAGGATGGGCTGTCGCTCCCCTTCGGCTGGACCCCCAGCAG 22622

    195 eLeuTyrAsnGlyPheProLeuProProValArgGluValThrArgHisV 212
              |||||||||   ||||||   ||||||   |||:::::::::::
  22623 GAGGCGCAATGGCTTCCTTCTCCCTCTTGTCCGGGCTGTCTCCAACCAGA 22672

    212 alIleGlnValSerAsnGluValValThrAspAspAspArgTyrSerAsp 228
        ::::::::      ||||||   :::|||:::|||         :::
  22673 TTGTGCGCTTCCCCAATGAGAGACTGACCTCCGACCGTGGCCGAGCCCTC 22722

    229 LeuLeuMetAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrPr 245
        ::::::|||   |||||||||:::||||||||||||:::   |||:::||
  22723 ATGTTCATGCAGTGGGGCCAGTTCATTGACCATGACCTGGACTTCTCCCC 22772

    245 oGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetT 262
        |:::|||   :::::::::||||||   :::|||   ||||||:::   |
  22773 GGAGTCCCCGGCCAGAGTGGCCTTCACTGCAGGCGTTGACTGTGAGAGGA 22822

    262 hrCysGluAsnGlnAsnProCysPheProIleGlnLeu...ProGluGlu 277
        |||||   :::      ||||||||||||||:::::   |||::::::
  22823 CCTGCGCCCAGCTGCCCCCCTGCTTTCCCATCAAGATCCCACCCAATGAC 22872
```

```
  278 AlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArgSerSerAl 294
      |||                  |||:::|||||||:::|||||||:::
22873 CCCCGCATCAAGAACCAGCGTGACTGCATCCCTTTCTTCCGCTCGGCACC 22922

  294 aAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrA 311
       :::|||      :::   :::
22923 CTCATGCCCCCAAAACAAGAAC..........................  22944

  311 laAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSer 327
        :::   |||:::|||:::|||:::||||||||||||:::|||||||||
22945 ..AGAGTCCGCAACCAGATCAACGCGCTCACCTCCTTTGTGGACGCCAGC 22992

  328 ThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAsnTrpTh 344
      ||||||||||||:::   :::|||:::   :::|||||||||   ||
22993 ATGGTGTATGGCAGTGAGGTCTCCCTCTCGCTGCGGCTCCGCAACCGGAC 23042

  344 rSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAspSerGlyA 361
      |:::      |||||||||   ::::::   |||:::::::|||:::||||
23043 CAACTACCTGGGGCTGCTGGCCATCAACCAGCGCTTTCAAGACAACGGCC 23092

  361 rgAlaTyrLeuProPheValProProArgAlaProAlaAlaCysAlaPro 377
      |||||   |||||||||   |||
23093 GGGCCCTGCTGCCCTTCGACCCCGA.......................... 23117

  378 GluProGlyAsnProGlyGluThrArgGlyProCysPheLeuAlaGlyAs 394

23117 ..................................................  23117

  394 pGlyArgAlaSerGluValProSerLeuThrAlaLeuHisThrLeuTrpL 411
             ::::::|||:::|||:::|||:::|||:::||||||||||:::::
23118 .......TCAACGGAAACCCCCAAACTGGCAGCCATGCACACCCTCTTTA 23160

  411 euArgGluHisAsnArgLeuAlaAlaAlaLeuLysAlaLeuAsnAlaHis 427
      ::|||||||||||||||||||||:::   |||:::   ||||||   :::
23161 TGCGAGAGCACAACCGGCTGGCCACCGAGCTGAGACGCCTGAATCCCCGG 23210

  428 TrpSerAlaAspAlaValTyrGlnGluAlaArgLysValValGlyAlaLe 444
      |||:::::::|||   :::|||:::||||||||||||:::::||||||::
23211 TGGAATGGAGACAAACTGTACAATGAGGCTCGGAAGATCATGGGGGCCAT 23260

  444 uHisGlnIleIleThrLeuArgAspTyrIleProArgIleLeuGlyProG 461
      :   ||||||||||||   ||||||:::::::|||   :::||||||
23261 GGTCCAGATCATCACCTACCGAGACTTTCTGCCCCTGGTTCTGGGCAAGG 23310

  461 luAlaPheGlnGlnTyrValGlyProTyrGluGlyTyrAspSerThrAla 477
             ::::::   :::|||   |||:::||||||   |||::::::
23311 CCCGGGCCAGGAGAACCCTGGGGCACTACAGGGGGTACTGCTCCAATGTG 23360

  478 AsnProThrValSerAsnValPheSerThrAlaAlaPheArgPheGlyHi 494
      :::|||   |||:::||||||||||   |||   ||||||||||||||||
23361 GACCCACGGGTGGCCAATGTCTTC...ACCCTGGCCTTCCGCTTTGGCCA 23407

  494 sAlaThrIleHisProLeuValArgArgLeuAspAlaSerPheGlnGluH 511
      |:::   ::::::|||::::::   |||||||||:::::::::::::
23408 CACAATGCTCCAGCCCTTCATGTTCCGCTTGGACAGTCAGTACCGGGCCT 23457

  511 isProAspLeuProGlyLeuTrpLeuHisGlnAlaPhePheSerProTrp 527
                    :::   |||   :::||||||||||:::   |||
23458 CCGCACCCAACTCGCATGTCCCACTTAGCTCTGCCTTCTTTGCCAGCTGG 23507

  528 ThrLeuLeuArgGlyGlyGlyLeuAspProLeuIleArgGlyLeuLeuAl 544
         ::::::      ||||||:::||||||:::::||||||||:::||
23508 CGGATCGTGTATGAAGGGGGCATCGACCCCATCCTCCGGGGCCTCATGGC 23557

  544 aArgProAlaLysLeuGlnValGlnAspGlnLeuMetAsnGluGluLeuT 561
      |   ||||||||||||:::   ||||||   ::::::   :::||||||
23558 CACCCCTGCCAAGCTGAACCGTCAGGATGCCATGTTAGTGGATGAGCTCC 23607

  561 hrGluArgLeuPheValLeuSerAsnSerSerThrLeuAspLeuAlaSer 577
        :::||||||||||          :::          ||||||||||||:::
23608 GGGACCGGCTGTTTCGGCAAGTGAGGAGGATTGGGCTGGACCTGGCAGCT 23657

  578 IleAsnLeuGlnArgGlyArgAspHisGlyLeuProGlyTyrAsnGluTr 594
```

```
         :::|||:::||||||:::|||||||||||||||||||||||||||||   ||
23658 CTCAACATGCAACGAAGCCGGGACCACGGCCTTCCAGGGTACAATGCTTG 23707

  594 pArgGluPheCysGlyLeuProArgLeuGluThrProAlaAspLeuSerT 611
      ||||:::||||||||||||   :::   ::::::   |||:::||||||
23708 GAGGCGCTTCTGTGGGCTCTCCCAGCCCCGGAATTTGGCACAGCTTAGCC 23757

  611 hrAlaIleAlaSerArgSerValAlaAspLysIleLeuAspLeuTyrLys 627
        ::::::   ::::::::::::|||   |||:::|||:::||||||
23758 GGGTGCTGAAAAACCAGGACTTGGCAAGGAAGTTCCTGAATTTGTATGGA 23807

  628 HisProAspAsnIleAspValTrpLeuGlyGlyLeuAlaGluAsnPheLe 644
        ||||||||||||||:::|||:::|||:::::::||||||   :::||
23808 ACACCTGACAACATTGACATCTGGATTGGGGCCATCGCTGAGCCTCTTTT 23857

  644 uProArgAlaArgThrGlyProLeuPheAlaCysLeuIleGlyLysGlnM 661
      ||||   ||||||:::||||||||||:::||||||||||:::   :::|||:
23858 GCCGGGGGCTCGAGTGGGGCCTCTTCTGGCTTGTCTGTTCGAGAACCAGT 23907

  661 etLysAlaLeuArgAspGlyAspTrpPheTrpTrpGluAsnSerHisVal 677
      :::::     ||||||||||||   |||||||||:::|||   :::|||
23908 TCAGAAGAGC.CGAGACGGAGACAGGTTCTGGTGGCAGAAC...GAGGTG 23953

  678 PheThr.AspAlaGlnArgArgGluLeuGluLysHisSerLeuSerArgV 694
      |||::: :::   ||||||:::   |||::::::   ||||||||||||:
23954 TTTTCACCAAAGACAGCGCAAGGCCCTGAGCAGAATTTCCTTGTCTCGAA 24003

  694 alIleCysAspAsnThrGlyLeuThrArgValProMetAspAlaPheGln 710
      ::|||||||||||||||||:::|||   |||       |||   |||:::
24004 TTATATGTGACAATACCGGTATCACCACGGTTTCAAGGGACATCTTCAGA 24053

  711 ValGlyLysPheProGluAspPheGluSerCysAspSerIleThrGlyMe 727
      ::::::   :::|||:::   |||   :::|||:::   |||      ::
24054 GCCAACATCTACCCTCGGGGCTTTGTGAACTGCAGCCGTATCCCCAGGTT 24103

  727 tAsnLeuGluAlaTrpArgGluThr 735
      :||||||:::|||||||||   |||
24104 GAACCTATCAGCCTGGCGAGGGACA 24128
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ51671

seq_documentation_block:
ID   AAZ51671 standard; cDNA; 5510 BP.
XX
AC   AAZ51671;
XX
DT   21-JUN-2000  (first entry)
XX
DE   Human p53 target molecule, PRG2 cDNA.
XX
KW   PRG2; p53 target; human; modulate; cell proliferation; immunomodulatory;
KW   chromosome 2p24.3; cytostatic; gene therapy; tumour cell; inducer;
KW   diagnosis; therapeutic; proliferative disease; cell cycle arrest; cancer;
KW   treatment; apoptosis; knockout animal; cancer susceptibility; dPxn; hPxn;
KW   peroxidasin; redox-regulation; reactive oxygen species; ROS; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4491
FT                   /*tag=  a
FT                   /product= "Human PRG2 protein"
FT                   /note= "Human homologue (hPxn) of Drospohila peroxidasin
FT                   gene, dPxn"
FT                   /partial
FT   sig_peptide     1..87
FT                   /*tag=  b
FT   mat_peptide     88..4488
FT                   /*tag=  c
FT                   /product= "Mature human PRG2 protein"
FT                   /note= "Exhibit growth inhibitory effect upon over
FT                   expression"
XX
```

```
PN   WO200012526-A1.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1999;   99WO-US19551.
XX
PR   28-AUG-1998;   98US-0098251.
XX
PA   (UYPR-) UNIV PRINCETON.
XX
PI   Horikoshi N,  Shenk T;
XX
DR   WPI; 2000-246724/21.
DR   P-PSDB; AAY70469.
XX
PT   New p53-inducible isolated nucleic acid molecule including open reading
PT   frame encoding human homolog of Drosophila melanogaster peroxidasin,
PT   useful e.g. in detection and treatment of cancer -
XX
PS   Disclosure; Page 68-69; 83pp; English.
XX
CC   The present sequence is the complete cDNA of PRG2 gene, whose RNA levels
CC   are upregulated in response to induction of p53 activity in human colon
CC   cancer EB1 cells. This sequence is the human homologue (hPxn) of
CC   Drosophila peroxidasin gene dPxn, that is expressed in heart, placenta,
CC   spleen, ovary and intestines. PRG2 is involved in p53-mediated growth
CC   suppression pathways and plays a role in redox regulation. It is a
CC   haem-peroxidase that increases the intracellular content of reactive
CC   oxygen species (ROS). They are potential targets of p53 regulatory
CC   activity and are useful for modulation of cellular proliferation. PRG2
CC   gene is localised to human chromosome 2p24.3. The PRG target molecules
CC   have cytostatic and immunomodulatory activity. PRG polynucleotides,
CC   proteins and antibodies are useful as diagnostic and therapeutic agents
CC   for detection and treatment of cancer and other proliferative diseases.
CC   The gene/cDNA may be used for gene therapy, to restore a gene function
CC   downstream of p53, that cannot be activated in the p53-deficient tumour
CC   cell. Antibodies can be used as inducers of cell cycle arrest and/or
CC   apoptosis. The DNA sequences can be used to generate 'knockout' animals
CC   as a model of cancer susceptibility.
XX
SQ   Sequence 5510 BP; 1258 A; 1638 C; 1527 G; 1087 T; 0 other;


alignment_scores:
           Quality: 1421.00              Length:    792
             Ratio:  2.671                Gaps:     21
 Percent Similarity:  67.172   Percent Identity: 39.899

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAZ51671   ..

 Align seg 1/1  to: AAZ51671  from: 1  to: 5510

       8 SerValThrLeuValMetAlaCysThrGluAlaPhePheProPheIleSe 24
         ||||||:::::::|||::::::   :::            |||   :::||
    1861 TCGGTGAGCATGGTGCTCAGTGTGAAC.........GTTCCTGACGTCAG 1901

      24 rArgGlyLysGluLeuLeuTrpGlyLysProGluGluSerArgValSerS 41
         ||||:::               |||   |||            |||::::
    1902 TCGAAAT..............GGAGATCCG.........TTTGTAGCTA 1927

      41 erValLeuGluGluSerLysArgLeuValAspThrAlaMetTyrAlaThr 57
         ::   :::   |||:::          ||||||   |||:::   :::|||
    1928 CCTCCATCGTGGAAGCGATTGCGACTGTTGACAGAGCTATAAACTCAACC 1977

      58 MetGlnArgAsnLeuLysLysArgGlyIleLeuSerGlyAlaGlnLeuLe 74
               :::   :::   :::|||         |||      :::|||||
    1978 CGAACACATTTGTTTGACAGCCGTCCT...CGTTCTCCAAATGATTTGCT 2024

      74 uSerPheSerLysLeuPro...GluProThrSerGlyValIleAlaArgA 90
         |:::::   :::   |||   :::|||   :::         |||||||
    2025 GGCCTTGTTCCGGTATCCGAGGGATCCTTACACAGTTGAACAGGCACGGG 2074

      90 laAlaGluIleMetGluThrSerIleGlnAlaMetLysArgLysValAsn 106
         ||:::||||||:::|||   :::::::|||   :::::::::   |||:::
    2075 CGGGAGAAATCTTTGAACGGACATTGCAGCTCATTCAGGAGCATGTACAG 2124

     107 LeuLysThrGln..............GlnSerGlnHisProThrAspAl 118
                 :::                 |||   |||   :::|||
    2125 CATGGCTTGATGGTCGACCTCAACGGAACAAGTTACCACTACAACGACCT 2174

     118 aLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysLeuP 135
          :::|||   :::   |||::::::|||||||||:::|||||||||
    2175 GGTGTCTCCACAGTACCTGAACCTCATCGCAAACCTGTCGGGCTGTACCG 2224

     135 roTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLysTyr 151
           :::            :::|||   :::   |||:::   :::||||||
    2225 CCCACCGGCGCGTGAACAACTGCTCGGACATGTGCTTCCACCAGAAGTAC 2274

     152 ArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGlyAl 168
         |||          |||:::|||||||||   :::||||||   ||||||||
    2275 CGGACGCACGACGGCACCTGTAACAACCTGCAGCACCCCATGTGGGGCGC 2324

     168 aSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspGlyP 185
         ||||   ||||||:::   |||   |||      ||||||||||:::||||
    2325 CTCGCTGACCGCCTTCGAGCGCCTGCTGAAATCCGTGTACGAGAATGGCT 2374

     185 heSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhePro 201
         ||:::   |||||||||   ||||||      ||||||||||
    2375 TCAACACCCCTCGGGGCATCAACCCCCACCGACTGTACAACGGGCACGCC 2424

     202 LeuProProValArgGluValThrArgHisValIleGlnValSerAsnGl 218
         ||||||      |||   |||:::      :::|||      :::::||
    2425 CTTCCCATGCCGCGCCTGGTGTCCACCACCCTGATC......GGGACGGA 2468

     218 uValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpGlyG 235
         |:::||||||   |||:::::::::::::   :::||||||   |||||||
    2469 GACCGTCACACCCGACGAGCAGTTCACCCACATGCTGATGCAGTGGGGCC 2518

     235 lnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLysAla 251
         ||:::::|||||||||:::      |||      :::   |||:::|||
    2519 AGTTCCTGGACCACGACCTCGACTCCACGGTGGTGGCCCTGAGCCAGGCA 2568

     252 AlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAsnPr 268
            |||:::   |||:::   |||:::   :::|||:::|||:::   ||
    2569 CGCTTCTCCGACGGACAGCACTGCAGCAACGTGTGCAGCAACGACCCCCC 2618

     268 oCysPheProIleGlnLeu...ProGluGluAlaArgProAlaAlaGlyT 284
         |||||||   :::::::::   |||::::::::::|||      :::|||:
    2619 CTGCTTCTCTGTCATGATCCCCCCCAATGACTCCCGGGCCAGGAGCGGGG 2668

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ::   |||:::   |||   |||||||||   :::|||||||:::|||
    2669 CCCGCTGCATGTTCTTCGTGCGCTCCAGCCCTGTGTGCGGCAGCGGCATG 2718

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
         :::   |||         ::::::::::::   ||||||:::|||::
    2719 ACTTCGCTGCTC.........ATGAACTCCGTGTACCCGCGGGAGCAGAT 2759

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         :|||   |||||||||:::::|||||||||:::|||||||||||:::
    2760 CAACCAGCTCACCTCCTACATCGACGCATCCAACGTGTACGGGAGCACGG 2809

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
                 |||:::::::|||:::   :::|||   :::||||||||||
    2810 AGCATGAGGCCCGCAGCATCCGCGACCTGGCCAGCCACCGCGGCCTGCTG 2859

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         |||   :::|||   ::::::   |||||||:::      |||||||||::
    2860 CGG...CAGGGCATCGTGCAGCGGTCCGGGAAGCCGCTGCTCCCCTTCGC 2906

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         :            |||:::   |||      :::            :::|
    2907 C...ACCGGGCCGCCCACGGAGTGCATGCGGGAC.........GAGAACG 2944

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
```

```
2945 AGAGCCCCATCCCCTGCTTCCTGGCCGGGGACCACCGCGCCAACGAGCAG 2994

 401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
2995 CTGGGCCTGACCAGCATGCACACGCTGTGGTTCCGCGAGCACAACCGCAT 3044

 417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
3045 TGCCACGGAGCTGCTCAAGCTGAACCCGCACTGGGACGGCGACACCATCT 3094

 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
3095 ACTATGAGACCAGGAAGATCGTGGGTGCGGAGATCCAGCACATCACCTAC 3144

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
3145 CAGCACTGGCTCCCGAAGATCCTGGGG...GAGGTGGGCATGAGGACGCT 3191

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
3192 GGGAGAGTACCACGGCTACGACCCCGGCATCAATGCTGGCATCTTCAACG 3241

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
3242 CCTTCGCCACCGCGGCCTTCAGGTTTGGCCACACGCTTGTCAACCCACTG 3291

 501 ValArgArgLeuAspAlaSerPhe.....GlnGluHisProAs 513
3292 CTTTACCGGCTGGACGAGAACTTCCAGCCCATTGCACAAGATCAC..... 3336

 513 pLeuProGlyLeuTrpLeuHisGlnAlaPhePheSerProTrpThrLeuL 530
3337 .CTCCCC.....CTTCACAAAGCTTTCTTCTCTCCCTTCCGGATTG 3376

 530 euArgGlyGlyGlyLeuAspProLeuIleArgGlyLeuLeuAlaArgPro 546
3377 TGAATGAGGGCGGCATCGATCCGCTTCTCAGGGGGCTGTTCGGGGTGGCG 3426

 547 AlaLysLeuGlnValGlnAspGlnLeuMetAsnGluGluLeuThrGluAr 563
3427 GGGAAAATGCGTGTGCCCTCGCAGCTGCTGAACACGGAGCTCACGGAGCG 3476

 563 gLeuPheValLeuSerAsnSerSerThrLeuAspLeuAlaSerIleAsnL 580
3477 GCTGTTCTCCATGGCACACACGGTGGCTCTGGACCTGGCGGCCATCAACA 3526

 580 euGlnArgGlyArgAspHisGlyLeuProGlyTyrAsnGluTrpArgGlu 596
3527 TCCAGCGGGGCCGGGACCACGGGATCCCACCCTACCACGACTACAGGGTC 3576

 597 PheCysGlyLeuProArgLeuGluThrProAlaAspLeuSerThrAlaIl 613
3577 TACTGCAATCTATCGGCGGCACACACGTTCGAGGACCTGAAAAATGAGAT 3626

 613 eAlaSerArgSerValAlaAspLysIleLeuAspLeuTyrLysHisProA 630
3627 TAAAAACCCTGAGATCCGGGAGAAACTGAAAAGGTTGTATGGCTCGACAC 3676

 630 spAsnIleAspValTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArg 646
3677 TCAACATCGACCTGTTTCCGGCGCTCGTGGTGGAGGACCTGGTGCCTGGC 3726

 647 AlaArgThrGlyProLeuPheAlaCysLeuIleGlyLysGlnMetLysAl 663
3727 AGCCGGCTGGGCCCCACCCTGATGTGTCTTCTCAGCACACAGTTCAAGCG 3776

 663 aLeuArgAspGlyAspTrpPheTrpTrpGluAsnSerHisValPheThrA 680
3777 CCTGCGAGATGGGGACAGGTTGTGGTATGAGAACCCTGGGGTGTTCTCCC 3826

 680 spAlaGlnArgArgGluLeuGluLysHisSerLeuSerArgValIleCys 696
```

```
3827 CGGCCCAGCTGACTCAGATCAAGCAGACGTCGCTGGCCAGGATCCTATGC 3876

 697 AspAsnThr...GlyLeuThrArgValProMetAspAlaPheGlnValGl 712
3877 GACAACGCGGACAACATCACCCGGGTGCAGAGCGACGTGTTCAGGGTGGC 3926

 712 yLysPheProGluAspPheGluSerCysAspSerIleThrGlyMetAsnL 729
3927 GGAGTTCCCTCACGGCTACGGCAGCTGTGACGAGATCCCCAGGGTGGACC 3976

 729 euGluAlaTrpArgGluThrPheProGlnAspAspLysCysGlyPhePro 745
3977 TCCGGGTGTGGCAGGAC.....TGC..... 3996

 746 GluSerValGluAsnGlyAspPheValHisCysGluGluSerGlyArgAr 762
3997 .....TGTGAAGAC..... 4005

 762 gValLeuValTyrSerCysArg.....HisGlyT 772
4006 .....TGTAGGACCAGGGGGCAGTTCAATGCCTTTTCCT 4039

 772 yrGluLeuGlnGlyArgGluGlnLeu 780
4040 ATCATTTCCGAGGCAGACGGTCTCTT 4065
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV99922

seq_documentation_block:

ID AAV99922 standard; cDNA; 6847 BP.
XX
AC AAV99922;
XX
DT 10-MAY-1999 (first entry)
XX
DE Melanoma associated antigen MG50 gene.
XX
KW MG50; melanoma gene-50; melanoma associated antigen; human;
KW T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;
KW therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..4491
FT /*tag= a
FT polyA_signal 6805..6810
FT /*tag= b
XX
PN WO9855133-A1.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-US11533.
XX
PR 06-JUN-1997; 97US-0870941.
XX
PA (REGC) UNIV CALIFORNIA.
PA (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
PI Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;
XX
DR WPI; 1999-080820/07.
DR P-PSDB; AAW81030.
XX
PT New MG50 melanoma associated antigen fragments - used to develop
PT products for the detection, treatment and prevention of
PT MG50-expressing cancers, e.g. melanoma, lung cancer or
PT rhabdomyosarcoma
XX
PS Claim 8; Page 38-45; 79pp; English.
XX
CC This is the nucleotide sequence of cDNA encoding a portion (see

```
CC   AAW81030) of the new human melanoma associated antigen MG50.
CC   Subtractive hybridisation of cDNA obtained from melanoma cell line
CC   MSM M-1 against an excess of mRNA from squamous lung carcinoma cell
CC   line Lu-1 was used to clone cDNA sequences differentially expressed
CC   in MSM M-1 cells.  12 Candidate clones were obtained, 6 of which
CC   were considered novel.  1 Clone, designated melanoma gene-50
CC   (MG50), was selected for further characterisation.  The 5' region
CC   of the gene has yet to be obtained.  MG50 mRNA was detected in
CC   melanoma, lung carcinoma and rhabdomyosarcoma cells, foetal brain,
CC   foetal heart and human placenta.  The invention also provides T
CC   cell epitopes (see AAW81031-54) from MG50, including cytotoxic and
CC   helper T cell epitopes, antibodies that specifically bind to MG50
CC   or an MG50 T cell epitope, recombinant vectors, and antigen
CC   presenting cells.  Methods are provided for identifying an MG50
CC   melanoma associated antigen in an individual and for identifying an
CC   immune response against an MG50 melanoma associated antigen, as
CC   well as methods of stimulating T lymphocytes that are reactive
CC   against cancer cells expressing MG50 and for treating an individual
CC   having cancer cells that express MG50.  The products and methods
CC   can be used for the detection, treatment and prevention of
CC   MG50-expressing cancers, e.g. melanomas, lung cancer or
CC   rhabdomyosarcoma.
XX
SQ   Sequence 6847 BP; 1614 A; 1914 C; 1856 G; 1451 T; 12 other;


alignment_scores:
           Quality: 1421.00              Length:    792
             Ratio:  2.671                Gaps:     21
 Percent Similarity:  67.172   Percent Identity: 39.899

alignment_block:
 US-08-482-402A-3_COPY_1_848 x AAV99922  ..

 Align seg 1/1  to: AAV99922  from: 1  to: 6847

       8 SerValThrLeuValMetAlaCysThrGluAlaPhePheProPheIleSe 24
         ||||||:::::|||:::::   :::            |||   :::||
    1861 TCGGTGAGCATGGTGCTCAGTGTGAAC.........GTTCCTGACGTCAG 1901

      24 rArgGlyLysGluLeuLeuTrpGlyLysProGluGluSerArgValSerS 41
         ||||:::              |||   |||           |||::::
    1902 TCGAAAT..............GGAGATCCG.........TTTGTAGCTA 1927

      41 erValLeuGluGluSerLysArgLeuValAspThrAlaMetTyrAlaThr 57
         ::   :::   |||:::       ||||||   |||:::   :::|||
    1928 CCTCCATCGTGGAAGCGATTGCGACTGTTGACAGAGCTATAAACTCAACC 1977

      58 MetGlnArgAsnLeuLysLysArgGlyIleLeuSerGlyAlaGlnLeuLe 74
            :::   :::   :::|||         |||      :::||||||
    1978 CGAACACATTTGTTTGACAGCCGTCCT...CGTTCTCCAAATGATTTGCT 2024

      74 uSerPheSerLysLeuPro...GluProThrSerGlyValIleAlaArgA 90
         |::::::   :::   |||   :::|||   :::         |||||||
    2025 GGCCTTGTTCCGGTATCCGAGGGATCCTTACACAGTTGAACAGGCACGGG 2074

      90 laAlaGluIleMetGluThrSerIleGlnAlaMetLysArgLysValAsn 106
         ||:::||||||:::|||   :::::::|||   :::::::::   |||:::
    2075 CGGGAGAAATCTTTGAACGGACATTGCAGCTCATTCAGGAGCATGTACAG 2124

     107 LeuLysThrGln..............GlnSerGlnHisProThrAspAl 118
                 :::                  |||   |||   :::|||
    2125 CATGGCTTGATGGTCGACCTCAACGGAACAAGTTACCACTACAACGACCT 2174

     118 aLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysLeuP 135
          :::|||   :::   |||::::::|||||||||:::|||||||||
    2175 GGTGTCTCCACAGTACCTGAACCTCATCGCAAACCTGTCGGGCTGTACCG 2224

     135 roTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLysTyr 151
           :::            :::|||   :::   |||:::   :::||||||
    2225 CCCACCGGCGCGTGAACAACTGCTCGGACATGTGCTTCCACCAGAAGTAC 2274

     152 ArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGlyAl 168
         |||         |||:::|||||||||   :::||||||   ||||||||
    2275 CGGACGCACGACGGCACCTGTAACAACCTGCAGCACCCCATGTGGGGCGC 2324

     168 aSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspGlyP 185
         ||||   ||||||:::   |||   |||      |||||||||:::||||
    2325 CTCGCTGACCGCCTTCGAGCGCCTGCTGAAATCCGTGTACGAGAATGGCT 2374

     185 heSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhePro 201
         ||:::   |||||||||   ||||||      ||||||||||||
    2375 TCAACACCCCTCGGGGCATCAACCCCCACCGACTGTACAACGGGCACGCC 2424

     202 LeuProProValArgGluValThrArgHisValIleGlnValSerAsnGl 218
         ||||||      |||   |||:::     :::|||      ::::::||
    2425 CTTCCCATGCCGCGCCTGGTGTCCACCACCCTGATC......GGGACGGA 2468

     218 uValValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpGlyG 235
         |:::||||||   |||:::::::::::::   :::||||||   |||||||
    2469 GACCGTCACACCCGACGAGCAGTTCACCCACATGCTGATGCAGTGGGGCC 2518

     235 lnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLysAla 251
         ||:::::::|||||||||:::      |||      :::   |||:::|||
    2519 AGTTCCTGGACCACGACCTCGACTCCACGGTGGTGGCCCTGAGCCAGGCA 2568

     252 AlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAsnPr 268
            |||:::   |||:::   |||:::   :::|||:::|||:::   ||
    2569 CGCTTCTCCGACGGACAGCACTGCAGCAACGTGTGCAGCAACGACCCCCC 2618

     268 oCysPheProIleGlnLeu...ProGluGluAlaArgProAlaAlaGlyT 284
         |||||||   :::::::::   |||::::::::|||      :::||||:
    2619 CTGCTTCTCTGTCATGATCCCCCCCAATGACTCCCGGGCCAGGAGCGGGG 2668

     284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
         ::   |||:::   |||   |||||||||   :::||||||:::|||
    2669 CCCGCTGCATGTTCTTCGTGCGCTCCAGCCCTGTGTGCGGCAGCGGCATG 2718

     301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
            :::   |||         ::::::::::::   ||||||:::|||::
    2719 ACTTCGCTGCTC.........ATGAACTCCGTGTACCCGCGGGAGCAGAT 2759

     317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
         :|||   |||||||||:::::|||||||||:::|||||||||||:::
    2760 CAACCAGCTCACCTCCTACATCGACGCATCCAACGTGTACGGGAGCACGG 2809

     334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
                 |||::::::|||:::   :::|||   :::|||||||||
    2810 AGCATGAGGCCCGCAGCATCCGCGACCTGGCCAGCCACCGCGGCCTGCTG 2859

     351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
         |||   :::|||   ::::::   ||||||:::      |||||||||::
    2860 CGG...CAGGGCATCGTGCAGCGGTCCGGGAAGCCGCTGCTCCCCTTCGC 2906

     367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
         :            |||:::   |||      :::            :::|
    2907 C...ACCGGGCCGCCCACGGAGTGCATGCGGGAC.........GAGAACG 2944

     384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
         ||:::      ||||||||||||||||||||   ||||||:::|||
    2945 AGAGCCCCATCCCCTGCTTCCTGGCCGGGGACCACCGCGCCAACGAGCAG 2994

     401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
            :::||||||::::::|||||||||||||:::|||||||||||||||::
    2995 CTGGGCCTGACCAGCATGCACACGCTGTGGTTCCGCGAGCACAACCGCAT 3044

     417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
         :|||:::   |||      ||||||   ||||||:::::|||::::::|
    3045 TGCCACGGAGCTGCTCAAGCTGAACCCGCACTGGGACGGCGACACCATCT 3094

     434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
         ||   |||:::||||||:::|||||||||      |||   ||||||
    3095 ACTATGAGACCAGGAAGATCGTGGGTGCGGAGATCCAGCACATCACCTAC 3144

     451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
         :::   ::::::|||:::|||||||||   |||:::   ::::::   ::
```

```
3145 CAGCACTGGCTCCCGAAGATCCTGGGG...GAGGTGGGCATGAGGACGCT 3191

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
     :|||   |||:::|||||||||         |||      :::   |||:
3192 GGGAGAGTACCACGGCTACGACCCCGGCATCAATGCTGGCATCTTCAACG 3241

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
     ::|||:::||||||||||||||||||||||||:::   ::::::||||||
3242 CCTTCGCCACCGCGGCCTTCAGGTTTGGCCACACGCTTGTCAACCCACTG 3291

 501 ValArgArgLeuAspAlaSerPhe.............GlnGluHisProAs 513
     :::   |||||||||   :::|||             |||:::|||
3292 CTTTACCGGCTGGACGAGAACTTCCAGCCCATTGCACAAGATCAC..... 3336

 513 pLeuProGlyLeuTrpLeuHisGlnAlaPhePheSerProTrpThrLeuL 530
      ||||||         ||||||:::|||||||||||||||:::   ::::
3337 .CTCCCC.........CTTCACAAAGCTTTCTTCTCTCCCTTCCGGATTG 3376

 530 euArgGlyGlyGlyLeuAspProLeuIleArgGlyLeuLeuAlaArgPro 546
     :::::   ||||||:::|||||||||:::|||||||||:::::
3377 TGAATGAGGGCGGCATCGATCCGCTTCTCAGGGGGCTGTTCGGGGTGGCG 3426

 547 AlaLysLeuGlnValGlnAspGlnLeuMetAsnGluGluLeuThrGluAr 563
     :::|||:::::|||   :::||||||:::|||   ||||||||||||||
3427 GGGAAAATGCGTGTGCCCTCGCAGCTGCTGAACACGGAGCTCACGGAGCG 3476

 563 gLeuPheValLeuSerAsnSerSerThrLeuAspLeuAlaSerIleAsnL 580
     |||||||   ::::::::::::   :::||||||||||||:::||||||:
3477 GCTGTTCTCCATGGCACACACGGTGGCTCTGGACCTGGCGGCCATCAACA 3526

 580 euGlnArgGlyArgAspHisGlyLeuProGlyTyrAsnGluTrpArgGlu 596
     ::|||||||||||||||||||:::|||   |||:::::::::|||
3527 TCCAGCGGGGCCGGGACCACGGGATCCCACCCTACCACGACTACAGGGTC 3576

 597 PheCysGlyLeuProArgLeuGluThrProAlaAspLeuSerThrAlaIl 613
     :::|||:::|||         :::|||      ||||||::::::   ||
3577 TACTGCAATCTATCGGCGGCACACACGTTCGAGGACCTGAAAAATGAGAT 3626

 613 eAlaSerArgSerValAlaAspLysIleLeuAspLeuTyrLysHisProA 630
     |   :::   ::::::   :::|||:::      ||||||
3627 TAAAAACCCTGAGATCCGGGAGAAACTGAAAAGGTTGTATGGCTCGACAC 3676

 630 spAsnIleAspValTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArg 646
       |||||||||::::::   :::   ::::::|||:::::::::|||
3677 TCAACATCGACCTGTTTCCGGCGCTCGTGGTGGAGGACCTGGTGCCTGGC 3726

 647 AlaArgThrGlyProLeuPheAlaCysLeuIleGlyLysGlnMetLysAl 663
     :::|||   ||||||   :::   ||||||:::::   |||:::|||
3727 AGCCGGCTGGGCCCCACCCTGATGTGTCTTCTCAGCACACAGTTCAAGCG 3776

 663 aLeuArgAspGlyAspTrpPheTrpTrpGluAsnSerHisValPheThrA 680
      |||||||||||||   :::|||:::||||||      ||||||:::
3777 CCTGCGAGATGGGGACAGGTTGTGGTATGAGAACCCTGGGGTGTTCTCCC 3826

 680 spAlaGlnArgArgGluLeuGluLysHisSerLeuSerArgValIleCys 696
       ||||||      ::::::::::::   ||||||:::|||:::::|||
3827 CGGCCCAGCTGACTCAGATCAAGCAGACGTCGCTGGCCAGGATCCTATGC 3876

 697 AspAsnThr...GlyLeuThrArgValProMetAspAlaPheGlnValGl 712
     ||||||:::   ::::::|||||||||      |||:::|||:::|||::
3877 GACAACGCGGACAACATCACCCGGGTGCAGAGCGACGTGTTCAGGGTGGC 3926

 712 yLysProProGluAspPheGluSerCysAspSerIleThrGlyMetAsnL 729
     ::::||||||:::   :::   |||||||||:::|||      ::::::|
3927 GGAGTTCCCTCACGGCTACGGCAGCTGTGACGAGATCCCCAGGGTGGACC 3976

 729 euGluAlaTrpArgGluThrProProGlnAspAspLysCysGlyPhePro 745
     ||::::::|||::::::                      |||
3977 TCCGGGTGTGGCAGGAC.....................TGC......... 3996

 746 GluSerValGluAsnGlyAspPheValHisCysGluGluSerGlyArgAr 762
                                 ||||||:::
3997 ............................TGTGAAGAC........... 4005

 762 gValLeuValTyrSerCysArg.....................HisGlyT 772
                      ||||||                      :::|
4006 ................TGTAGGACCAGGGGGCAGTTCAATGCCTTTTCCT 4039

 772 yrGluLeuGlnGlyArgGluGlnLeu 780
     ||:::::::::|||||:::::|||
4040 ATCATTTCCGAGGCAGACGGTCTCTT 4065
```